GenCore version 4.5

OM protein - protein search, using sw model

Run on: June 26, 2002, 12:36:27 ; Search time 67.61 Seconds
(without alignments)
199.783 Million cell updates/sec

Title: US-09-838-785-2
Perfect score: 2861
Sequence: 1 MVQRLWVSRLLRHRKAQLLL.....AIYFATQVVFDKSDLAKYSA 553

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2861	100.0	553	4	US-09-020-956-113	Sequence 113, App
2	2861	100.0	553	4	US-09-030-607-113	Sequence 113, App
3	2861	100.0	553	4	US-09-439-313-113	Sequence 113, App
4	1287	45.0	255	4	US-09-071-710-36	Sequence 36, Appl
5	1287	45.0	255	4	US-09-525-397-36	Sequence 36, Appl
6	452	15.8	84	4	US-09-439-313-571	Sequence 571, App
7	330	11.5	516	1	US-08-356-340-4	Sequence 4, Appli
8	330	11.5	516	2	US-08-786-555-4	Sequence 4, Appli
9	324.5	11.3	525	1	US-08-356-340-2	Sequence 2, Appli
10	324.5	11.3	525	2	US-08-786-555-2	Sequence 2, Appli
11	312	10.9	58	4	US-09-439-313-547	Sequence 547, App
12	304	10.6	56	4	US-09-439-313-564	Sequence 564, App
13	243	8.5	44	4	US-09-071-710-37	Sequence 37, Appl
14	243	8.5	44	4	US-09-525-397-37	Sequence 37, Appl
15	157	5.5	29	4	US-09-439-313-546	Sequence 546, App
16	146	5.1	27	4	US-09-071-710-39	Sequence 39, Appl
17	146	5.1	27	4	US-09-525-397-39	Sequence 39, Appl
18	146	5.1	27	4	US-09-439-313-566	Sequence 566, App
19	137	4.8	25	4	US-09-439-313-520	Sequence 520, App
20	135	4.7	27	4	US-09-071-710-38	Sequence 38, Appl
21	135	4.7	27	4	US-09-525-397-38	Sequence 38, Appl
22	120	4.2	21	4	US-09-439-313-496	Sequence 496, App
23	120	4.2	21	4	US-09-439-313-521	Sequence 521, App
24	119	4.2	22	4	US-09-439-313-565	Sequence 565, App
25	115	4.0	22	4	US-09-439-313-559	Sequence 559, App
26	114	4.0	22	4	US-09-439-313-557	Sequence 557, App
27	114	4.0	22	4	US-09-439-313-561	Sequence 561, App
28	112	3.9	20	4	US-09-439-313-495	Sequence 495, App
29	110	3.8	427	4	US-08-900-230-4	Sequence 4, Appli
30	109.5	3.8	455	1	US-08-035-928-2	Sequence 2, Appli
31	109	3.8	22	4	US-09-439-313-555	Sequence 555, App
32	107	3.7	20	4	US-09-439-313-493	Sequence 493, App
33	107	3.7	20	4	US-09-439-313-497	Sequence 497, App
34	106.5	3.7	436	6	5432081-10	Patent No. 5432081
35	106	3.7	22	4	US-09-439-313-569	Sequence 569, App
36	106	3.7	427	4	US-09-199-737-4	Sequence 4, Appli
37	105	3.7	22	4	US-09-439-313-553	Sequence 553, App
38	105	3.7	22	4	US-09-439-313-572	Sequence 572, App
39	101	3.5	18	4	US-09-439-313-548	Sequence 548, App
40	100	3.5	20	4	US-09-439-313-490	Sequence 490, App
41	100	3.5	20	4	US-09-439-313-499	Sequence 499, App
42	99	3.5	20	4	US-09-439-313-563	Sequence 563, App
43	99	3.5	20	4	US-09-439-313-570	Sequence 570, App
44	98	3.4	20	4	US-09-439-313-498	Sequence 498, App
45	98	3.4	713	3	US-09-335-409-11	Sequence 11, Appl

ALIGNMENTS

RESULT 1
US-09-020-956-113
; Sequence 113, Application US/09020956
; Patent No. 6261562
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillin, Davin C.
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE CANCER AND METHODS
; NUMBER OF SEQUENCES: 178
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/020,956
; FILING DATE: 09-FEB-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.427C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 113:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 553 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
US-09-020-956-113

Query Match 100.0%; Score 2861; DB 4; Length 553;
Best Local Similarity 100.0%; Pred. No. 3.7e-271;
Matches 553; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MVQRLWVSRLLRHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIG 60

```
Db      1 MVQRLWVSRLLRHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIG 60

Qy     61 PVLGLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDPRPL 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 PVLGLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDPRPL 120

Qy    121 ELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPA 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 ELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPA 180

Qy    181 IDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPH 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 IDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPH 240

Qy    241 CCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGL 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 CCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGL 300

Qy    301 YQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVA 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 YQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVA 360

Qy    361 AFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 AFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420

Qy    421 ASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 ASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480

Qy    481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFATQ 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFATQ 540

Qy    541 VVFDKSDLAKYSA 553
          |||||||||||||
Db    541 VVFDKSDLAKYSA 553


RESULT   2
US-09-030-607-113
; Sequence 113, Application US/09030607
; Patent No. 6262245
;  GENERAL INFORMATION:
;    APPLICANT:  Xu, Jiangchun
;    APPLICANT:  Dillon, Davin C.
;    TITLE OF INVENTION:  COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE CANCER AND METHODS FO
;    NUMBER OF SEQUENCES:  224
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  WA
;      COUNTRY:  USA
;      ZIP:  98104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/030,607
;      FILING DATE:  25-FEB-1998
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Maki, David J.
;      REGISTRATION NUMBER:  31,392
;      REFERENCE/DOCKET NUMBER:  210121.427C3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;  INFORMATION FOR SEQ ID NO:  113:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  553 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
US-09-030-607-113


  Query Match             100.0%;  Score 2861;  DB 4;  Length 553;
  Best Local Similarity   100.0%;  Pred. No. 3.7e-271;
  Matches  553;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 MVQRLWVSRLLRHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MVQRLWVSRLLRHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIG 60

Qy     61 PVLGLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDPRPL 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 PVLGLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDPRPL 120

Qy    121 ELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPA 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 ELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPA 180

Qy    181 IDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPH 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 IDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPH 240

Qy    241 CCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGL 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 CCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGL 300

Qy    301 YQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVA 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 YQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVA 360

Qy    361 AFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 AFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420

Qy    421 ASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 ASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480

Qy    481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFATQ 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFATQ 540

Qy    541 VVFDKSDLAKYSA 553
          |||||||||||||
Db    541 VVFDKSDLAKYSA 553


RESULT   3
US-09-439-313-113
; Sequence 113, Application US/09439313
; Patent No. 6329505
; GENERAL INFORMATION:
;  APPLICANT: Xu, Jiangchun
;  APPLICANT:  Dillon, Davin C.
;  APPLICANT:  Mitcham, Jennifer L.
;  APPLICANT:  Harlocker, Susan Louise
;  APPLICANT:  Jiang Yuqui
;  APPLICANT:  Reed, Steven G.
;  APPLICANT:  Kalos, Michael
;  APPLICANT:  Fanger, Gary
;  APPLICANT:  Retter, Mark
;  APPLICANT:  Solk, John
;  APPLICANT:  Day, Craig
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
;  TITLE OF INVENTION:  DIAGNOSIS OF PROSTATE CANCER
```

```
;  FILE REFERENCE: 210121.427C9
;  CURRENT APPLICATION NUMBER: US/09/439,313
;  CURRENT FILING DATE:  1999-11-12
;  NUMBER OF SEQ ID NOS: 575
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 113
;   LENGTH: 553
;   TYPE: PRT
;   ORGANISM: Homo sapien
US-09-439-313-113


  Query Match           100.0%;  Score 2861;  DB 4;  Length 553;
  Best Local Similarity  100.0%;  Pred. No. 3.7e-271;
  Matches  553;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 MVQRLWVSRLLRHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MVQRLWVSRLLRHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIG 60

Qy     61 PVLGLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDPRPL 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 PVLGLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDPRPL 120

Qy    121 ELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPA 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 ELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPA 180

Qy    181 IDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPH 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 IDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPH 240

Qy    241 CCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGL 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 CCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGL 300

Qy    301 YQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVA 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 YQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVA 360

Qy    361 AFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 AFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420

Qy    421 ASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 ASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480

Qy    481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFATQ 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFATQ 540

Qy    541 VVFDKSDLAKYSA 553
          |||||||||||||
Db    541 VVFDKSDLAKYSA 553


RESULT   4
US-09-071-710-36
; Sequence 36, Application US/09071710
; Patent No. 6130043
;  GENERAL INFORMATION:
;    APPLICANT:  BILLING-MEDEL, PATRICIA
;    APPLICANT:  COHEN, MAURICE
;    APPLICANT:  COLPITTS, TRACEY L.
;    APPLICANT:  FRIEDMAN, PAULA N.
;    APPLICANT:  GORDON, JULIAN
;    APPLICANT:  GRANADOS, EDWARD N.
;    APPLICANT:  HODGES, STEVEN C.
;    APPLICANT:  KLASS, MICHAEL R.
;    APPLICANT:  KRATOCHVIL, JON D.
;    APPLICANT:  ROBERTS-RAPP, LISA
;    APPLICANT:  RUSSELL, JOHN C.
;    APPLICANT:  STROUPE, STEPHEN D.
;    TITLE OF INVENTION:  REAGENTS AND METHODS USEFUL
;    TITLE OF INVENTION:  FOR DETECTING DISEASES OF THE PROSTATE
;    NUMBER OF SEQUENCES:  41
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Abbott Laboratories
;      STREET:  100 Abbott Park Road
;      CITY:  Abbott Park
;      STATE:  IL
;      COUNTRY:  USA
;      ZIP:  60064-3500
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/071,710
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/850,713
;      FILING DATE:  02-MAY-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Becker, Cheryl L.
;      REGISTRATION NUMBER:  35,441
;      REFERENCE/DOCKET NUMBER:  6083.US.P1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  847/935-1729
;      TELEFAX:  847/938-2623
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  36:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  255 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  No. 6130043e
US-09-071-710-36


  Query Match           45.0%;  Score 1287;  DB 4;  Length 255;
  Best Local Similarity  100.0%;  Pred. No. 8.1e-118;
  Matches  255;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    299 GLYQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLAS 358
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GLYQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLAS 60

Qy    359 VAAFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDT 418
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 VAAFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDT 120

Qy    419 GGASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGEPT 478
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 GGASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGEPT 180

Qy    479 EARVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFA 538
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 EARVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFA 240

Qy    539 TQVVFDKSDLAKYSA 553
          |||||||||||||||
Db    241 TQVVFDKSDLAKYSA 255


RESULT   5
US-09-525-397-36
; Sequence 36, Application US/09525397
; Patent No. 6252047
;  GENERAL INFORMATION:
;    APPLICANT:  BILLING-MEDEL, PATRICIA
;    APPLICANT:  COHEN, MAURICE
```

```
;    APPLICANT:  COLPITTS, TRACEY L.
;    APPLICANT:  FRIEDMAN, PAULA N.
;    APPLICANT:  GORDON, JULIAN
;    APPLICANT:  GRANADOS, EDWARD N.
;    APPLICANT:  HODGES, STEVEN C.
;    APPLICANT:  KLASS, MICHAEL R.
;    APPLICANT:  KRATOCHVIL, JON D.
;    APPLICANT:  ROBERTS-RAPP, LISA
;    APPLICANT:  RUSSELL, JOHN C.
;    APPLICANT:  STROUPE, STEPHEN D.
;    TITLE OF INVENTION:  REAGENTS AND METHODS USEFUL
;    TITLE OF INVENTION:  FOR DETECTING DISEASES OF THE PROSTATE
;    NUMBER OF SEQUENCES:  41
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Abbott Laboratories
;      STREET:  100 Abbott Park Road
;      CITY:  Abbott Park
;      STATE:  IL
;      COUNTRY:  USA
;      ZIP:  60064-3500
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/525,397
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  09/071,710
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Becker, Cheryl L.
;      REGISTRATION NUMBER:  35,441
;      REFERENCE/DOCKET NUMBER:  6083.US.P1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  847/935-1729
;      TELEFAX:  847/938-2623
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  36:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  255 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  No. 6252047e
US-09-525-397-36


  Query Match             45.0%;  Score 1287;  DB 4;  Length 255;
  Best Local Similarity   100.0%;  Pred. No. 8.1e-118;
  Matches  255;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     299 GLYQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLAS 358
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 GLYQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLAS 60

Qy     359 VAAFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDT 418
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 VAAFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDT 120

Qy     419 GGASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGEPT 478
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 GGASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGEPT 180

Qy     479 EARVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFA 538
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 EARVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFA 240

Qy     539 TQVVFDKSDLAKYSA 553
           |||||||||||||||
Db     241 TQVVFDKSDLAKYSA 255
```

```
RESULT   6
US-09-439-313-571
; Sequence 571, Application US/09439313
; Patent No. 6329505
; GENERAL INFORMATION:
;  APPLICANT: Xu, Jiangchun
;  APPLICANT:  Dillon, Davin C.
;  APPLICANT:  Mitcham, Jennifer L.
;  APPLICANT:  Harlocker, Susan Louise
;  APPLICANT:  Jiang Yuqui
;  APPLICANT:  Reed, Steven G.
;  APPLICANT:  Kalos, Michael
;  APPLICANT:  Fanger, Gary
;  APPLICANT:  Retter, Mark
;  APPLICANT:  Solk, John
;  APPLICANT:  Day, Craig
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
;  TITLE OF INVENTION:  DIAGNOSIS OF PROSTATE CANCER
;  FILE REFERENCE: 210121.427C9
;  CURRENT APPLICATION NUMBER: US/09/439,313
;  CURRENT FILING DATE:  1999-11-12
;  NUMBER OF SEQ ID NOS: 575
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 571
;   LENGTH: 84
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-439-313-571


  Query Match             15.8%;  Score 452;  DB 4;  Length 84;
  Best Local Similarity   100.0%;  Pred. No. 6.5e-37;
  Matches   84;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     404 HREKQVFLPKYRGDTGGASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGA 463
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 HREKQVFLPKYRGDTGGASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGA 60

Qy     464 SACDVSVRVVVGEPTEARVVPGRG 487
           ||||||||||||||||||||||||
Db      61 SACDVSVRVVVGEPTEARVVPGRG 84
```

```
RESULT   7
US-08-356-340-4
; Sequence 4, Application US/08356340
; Patent No. 5608146
;  GENERAL INFORMATION:
;    APPLICANT:  FROMMER, Wolf-Bernd
;    APPLICANT:  RIESMEIER, Jorg
;    TITLE OF INVENTION:  DNA SEQUENCES WITH OLIGOSACCHARIDE
;    TITLE OF INVENTION:  TRANSPORTER, PLASMIDS, BACTERIA AND PLANTS CONTAINING A
;    TITLE OF INVENTION:  TRANSPORTER AS WELL AS A PROCESS FOR THE PREPARATION AND
;    TITLE OF INVENTION:  TRANSFORMATION OF YEAST STRAINS FOR THE IDENTIFI...
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ostrolenk, Faber, Gerb & Soffen
;      STREET:  1180 Avenue of the Americas
;      CITY:  New York
;      STATE:  NY
;      COUNTRY:  US
;      ZIP:  10036-8403
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/356,340
;      FILING DATE:  21-DEC-1994
;      CLASSIFICATION:  800
```

```
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US PCT/EP93/01604
;      FILING DATE:  22-JUN-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  DE P42 20 759.2
;      FILING DATE:  24-JUN-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meilman, Edward A.
;      REGISTRATION NUMBER:  24,735
;      REFERENCE/DOCKET NUMBER:  P/951-106
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 382-0700
;      TELEFAX:  (212) 382-0888
;      TELEX:  236925
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  516 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-356-340-4


  Query Match             11.5%;  Score 330;  DB 1;  Length 516;
  Best Local Similarity   24.7%;  Pred. No. 7.4e-24;
  Matches  132;  Conservative   93;  Mismatches  197;  Indels  112;  Gaps   21;

Qy       4 RLWVSRLLRHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIGPVL 63
           :||          ::::|  :  |::   |  :: : | :  :|:  || : :   ||:
Db      29 KLW----------KIIVVASIAAGVQFGWALQLSLLTPYVQLLGIPHKFASFIWLCGPIS 78

Qy      64 GLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGL-------LCPD 116
             |::  |::|  ||:   |:|||||||| | :  :::::|||   |: | |       |
Db      79 GMIVQPVVGYYSDNCSSRFGRRRPFIAAGAALVMIAVFLI---GFAADLGHASGDTLGKG 135

Qy     117 PRPLELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPD-HCRQAYSVYAFMISLGGCLG 175
            :|  :|: ::|  :||    :   |  |||:||        | | : ::| :::|  ||
Db     136 FKPRAIAVFVVGFWILDVANNMLQGPCRALLADLSGGKSGRMRTANAFFSFFMAVGNILG 195

Qy     176 YLLPAIDW--------DTSALAPYLGTQEECLF-GLLTLIFLTCVAATLLVAEE-AALGP 225
           |   :            | |   |    : | |  :  |: || :| ||:   |
Db     196 YAAGSYSHLFKVFPFSKTKACDMYCANLKSCFFIAIFLLLSLTTIALTLVRENELPEKDE 255

Qy     226 TEPAEGLSAPSLSPHCCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMAL 285
           |  | |:           :::: |   | :   | :|    || :  | :    :|:|
Db     256 QEIDEKLAGAG--------KSKVPF--FGEIFGALKEL----PRPMWILLLVTCLNWIAW 301

Qy     286 MTFTLFYTDFVGEGLYQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRL 345
           | |: ||:: : :: |    : |    | || ||| |::|| ||  :    || :: |
Db     302 FPFFLYDTDWMAKEVFGG----QVGD--ARLYDLGVRAGAMGLLLQSVVLGFMSLGVEFL 355

Qy     346 VQRFGTRAVYLASVAAFPVAAGATCLSHSVAVV---------TASAALTGFT-------- 388
           :: |  |  |  :  | :   | ||: :: |           :  | | |
Db     356 GKKIG-GAKRLWGILNFVL---AICLAMTILVTKMAEKSRQHDPAGTLMGPTPGVKIGAL 411

Qy     389 --FSALQI-------LPYTLASLYHREK------------QVFLPKYRGDTGGASSEDS 426
             |:|| |       :|: |||::   :              : :|:      |   :|
Db     412 LLFAALGIPLAATFSIPFALASIFSSNRGSGQGLSLGVLNLAIVVPQMLVSLVGGPWDDL 471

Qy     427 LMTSFLPGPKPGAPFPNGHVGAGGSG-----LLPPPPALCGASACDVSVRVVVG 475
               |||      |  | | |  |      :|| |||       |    | :|
Db     472 FGGGNLPG------FVVGAVAAAASAVLALTMLPSPPA-------DAKPAVAMG 512


RESULT   8
US-08-786-555-4
; Sequence 4, Application US/08786555B
; Patent No. 5981181
; GENERAL INFORMATION:
;  APPLICANT: FROMMER, Wolf-Bernd
;  APPLICANT: RIESMEIER, Jorg
;  TITLE OF INVENTION: DNA SEQUENCES WITH OLIGOSACCHARIDE TRANSPORTER,
;  TITLE OF INVENTION: PLASMIDS, BACTERIA AND PLANTS CONTAINING A TRANSPORTER
;  TITLE OF INVENTION: AS WELL AS A PROCESS FOR PREPARATION AND TRANSFORMATION
;  TITLE OF INVENTION: OF YEAST STRAINS FOR THE IDENTIFICATION
;  FILE REFERENCE: Frommer
;  CURRENT APPLICATION NUMBER: US/08/786,555B
;  CURRENT FILING DATE: 1997-01-21
;  EARLIER APPLICATION NUMBER: 08/356,340
;  EARLIER FILING DATE: 1994-12-21
;  EARLIER APPLICATION NUMBER: PCT/EP93/01604
;  EARLIER FILING DATE: 1993-06-22
;  EARLIER APPLICATION NUMBER: DE P4220759.2
;  EARLIER FILING DATE: 1992-06-24
;  NUMBER OF SEQ ID NOS: 11
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
;   LENGTH: 516
;   TYPE: PRT
;   ORGANISM: Spinacia oleracea
US-08-786-555-4


  Query Match             11.5%;  Score 330;  DB 2;  Length 516;
  Best Local Similarity   24.7%;  Pred. No. 7.4e-24;
  Matches  132;  Conservative   93;  Mismatches  197;  Indels  112;  Gaps   21;

Qy       4 RLWVSRLLRHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIGPVL 63
           :||          ::::|  :  |::   |  :: : | :  :|:  || : :   ||:
Db      29 KLW----------KIIVVASIAAGVQFGWALQLSLLTPYVQLLGIPHKFASFIWLCGPIS 78

Qy      64 GLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGL-------LCPD 116
             |::  |::|  ||:   |:|||||||| | :  :::::|||   |: | |       |
Db      79 GMIVQPVVGYYSDNCSSRFGRRRPFIAAGAALVMIAVFLI---GFAADLGHASGDTLGKG 135

Qy     117 PRPLELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPD-HCRQAYSVYAFMISLGGCLG 175
            :|  :|: ::|  :||    :   |  |||:||        | | : ::| :::|  ||
Db     136 FKPRAIAVFVVGFWILDVANNMLQGPCRALLADLSGGKSGRMRTANAFFSFFMAVGNILG 195

Qy     176 YLLPAIDW--------DTSALAPYLGTQEECLF-GLLTLIFLTCVAATLLVAEE-AALGP 225
           |   :            | |   |    : | |  :  |: || :| ||:   |
Db     196 YAAGSYSHLFKVFPFSKTKACDMYCANLKSCFFIAIFLLLSLTTIALTLVRENELPEKDE 255

Qy     226 TEPAEGLSAPSLSPHCCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMAL 285
           |  | |:           :::: |   | :   | :|    || :  | :    :|:|
Db     256 QEIDEKLAGAG--------KSKVPF--FGEIFGALKEL----PRPMWILLLVTCLNWIAW 301

Qy     286 MTFTLFYTDFVGEGLYQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRL 345
           | |: ||:: : :: |    : |    | || ||| |::|| ||  :    || :: |
Db     302 FPFFLYDTDWMAKEVFGG----QVGD--ARLYDLGVRAGAMGLLLQSVVLGFMSLGVEFL 355

Qy     346 VQRFGTRAVYLASVAAFPVAAGATCLSHSVAVV---------TASAALTGFT-------- 388
           :: |  |  |  :  | :   | ||: :: |           :  | | |
Db     356 GKKIG-GAKRLWGILNFVL---AICLAMTILVTKMAEKSRQHDPAGTLMGPTPGVKIGAL 411

Qy     389 --FSALQI-------LPYTLASLYHREK------------QVFLPKYRGDTGGASSEDS 426
             |:|| |       :|: |||::   :              : :|:      |   :|
Db     412 LLFAALGIPLAATFSIPFALASIFSSNRGSGQGLSLGVLNLAIVVPQMLVSLVGGPWDDL 471

Qy     427 LMTSFLPGPKPGAPFPNGHVGAGGSG-----LLPPPPALCGASACDVSVRVVVG 475
               |||      |  | | |  |      :|| |||       |    | :|
Db     472 FGGGNLPG------FVVGAVAAAASAVLALTMLPSPPA-------DAKPAVAMG 512


RESULT   9
US-08-356-340-2
; Sequence 2, Application US/08356340
; Patent No. 5608146
;  GENERAL INFORMATION:
;    APPLICANT:  FROMMER, Wolf-Bernd
;    APPLICANT:  RIESMEIER, Jorg
;    TITLE OF INVENTION:  DNA SEQUENCES WITH OLIGOSACCHARIDE
;    TITLE OF INVENTION:  TRANSPORTER, PLASMIDS, BACTERIA AND PLANTS CONTAINING A
;    TITLE OF INVENTION:  TRANSPORTER AS WELL AS A PROCESS FOR THE PREPARATION AND
;    TITLE OF INVENTION:  TRANSFORMATION OF YEAST STRAINS FOR THE IDENTIFI...
;    NUMBER OF SEQUENCES:  4
```

```
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ostrolenk, Faber, Gerb & Soffen
;      STREET:  1180 Avenue of the Americas
;      CITY:  New York
;      STATE:  NY
;      COUNTRY:  US
;      ZIP:  10036-8403
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/356,340
;      FILING DATE:  21-DEC-1994
;      CLASSIFICATION:  800
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US PCT/EP93/01604
;      FILING DATE:  22-JUN-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  DE P42 20 759.2
;      FILING DATE:  24-JUN-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meilman, Edward A.
;      REGISTRATION NUMBER:  24,735
;      REFERENCE/DOCKET NUMBER:  P/951-106
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212) 382-0700
;      TELEFAX:  (212) 382-0888
;      TELEX:  236925
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  525 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-356-340-2


  Query Match             11.3%;  Score 324.5;  DB 1;  Length 525;
  Best Local Similarity   23.2%;  Pred. No. 2.6e-23;
  Matches  129;  Conservative   91;  Mismatches  196;  Indels  141;  Gaps   20;

Qy      23 LLTFGLEVCLAAGITY--------VPPLLLEVGVEEKFMTMVLGIGPVLGLVCVPLLGSA 74
           |   ||   :|||: :        : | :  :|:   :   :   ||: |::  ||:|
Db      35 LKKLGLVASVAAGVQFGWALQLSLLTPYVQLLGIPHTWAAYIWLCGPISGMIVQPLVGYY 94

Qy      75 SDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLA--GLLCPDP-----RPLELALLIL 127
           ||    |:|||||| | :  : ::: ||   |: | |    ||     :|  :|: ::
Db      95 SDRCTSRFGRRRPFIAAGAALVAVAVGLI---GFAADIGAASGDPTGNVAKPRAIAVFVV 151

Qy     128 GVGLLDFCGQVCFTPLEALLSDLFR-DPDHCRQAYSVYAFMISLGGCLGYLLPAID---- 182
           |  :||       |  |||:|:       | | : ::| ::||   ||   :
Db     152 GFWILDVANNTLQGPCRALLADMAAGSQTKTRYANAFFSFFMALGNIGGYAAGSYSRLYT 211

Qy     183 ----WDTSALAPYLGTQEECLFGLLT-LIFLTCVAATLLVAEEAALGPTEPAEGLSAPSL 237
               |:|   |    : | |  :| || || :| :::   :  :   :  | |   :
Db     212 VFPFTKTAACDVYCANLKSCFFISITLLIVLTILALSVVKERQITIDEIQEEEDLKNRNN 271

Qy     238 SPHCCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVG 297
           |  |    |||  |   | |:  |  |    |: :  | :     :|:|   | || ||::|
Db     272 SSGC----ARLPF--FGQLIGALKDL----PKPMLILLLVTALNWIAWFPFLLFDTDWMG 321

Qy     298 EGLYQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFG--TRAVY 355
           : :| |           : ||:||  |:||| :   :  | || :: | :  |   |
Db     322 KEVYGGT------VGEGKLYDQGVHAGALGLMINSVVLGVMSLSIEGLARMVGGAKRLWG 375

Qy     356 LASVAAFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYR 415
           : ::        | ||: :| :|| ||                    : |:    :
Db     376 IVNIIL------AVCLAMTV-LVTKSAE------------------HFRDSHHIM---- 405

Qy     416 GDTGGASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVG 475
                                          || : |||||  |     :::  |:|

Db     406 ---------------------------------GSAVPPPPPA--GVKGGALAIFAVLG 429

Qy     476 EPTEARV-VP----------GRGICLDLAILDSAFLLSQV----------------AP 506
             |     :|           | |  | | :|: | :: |:                 |
Db     430 IPLAITFSIPLALASIFSASSGSGQGLSLGVLNLAIVVPQMFVSVTSGPWDAMFGGGNLP 489

Qy     507 SLFMGSIVQLSQSVTAY 523
           :  :|::   : :| ::
Db     490 AFVVGAVAATASAVLSF 506


RESULT  10
US-08-786-555-2
; Sequence 2, Application US/08786555B
; Patent No. 5981181
; GENERAL INFORMATION:
;  APPLICANT: FROMMER, Wolf-Bernd
;  APPLICANT: RIESMEIER, Jorg
;  TITLE OF INVENTION: DNA SEQUENCES WITH OLIGOSACCHARIDE TRANSPORTER,
;  TITLE OF INVENTION: PLASMIDS, BACTERIA AND PLANTS CONTAINING A TRANSPORTER
;  TITLE OF INVENTION: AS WELL AS A PROCESS FOR PREPARATION AND TRANSFORMATION
;  TITLE OF INVENTION: OF YEAST STRAINS FOR THE IDENTIFICATION
;  FILE REFERENCE: Frommer
;  CURRENT APPLICATION NUMBER: US/08/786,555B
;  CURRENT FILING DATE: 1997-01-21
;  EARLIER APPLICATION NUMBER: 08/356,340
;  EARLIER FILING DATE: 1994-12-21
;  EARLIER APPLICATION NUMBER: PCT/EP93/01604
;  EARLIER FILING DATE: 1993-06-22
;  EARLIER APPLICATION NUMBER: DE P4220759.2
;  EARLIER FILING DATE: 1992-06-24
;  NUMBER OF SEQ ID NOS: 11
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 525
;   TYPE: PRT
;   ORGANISM: Spinacia oleracea
US-08-786-555-2


  Query Match             11.3%;  Score 324.5;  DB 2;  Length 525;
  Best Local Similarity   23.2%;  Pred. No. 2.6e-23;
  Matches  129;  Conservative   91;  Mismatches  196;  Indels  141;  Gaps   20;

Qy      23 LLTFGLEVCLAAGITY--------VPPLLLEVGVEEKFMTMVLGIGPVLGLVCVPLLGSA 74
           |   ||   :|||: :        : | :  :|:   :   :   ||: |::  ||:|
Db      35 LKKLGLVASVAAGVQFGWALQLSLLTPYVQLLGIPHTWAAYIWLCGPISGMIVQPLVGYY 94

Qy      75 SDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLA--GLLCPDP-----RPLELALLIL 127
           ||    |:|||||| | :  : ::: ||   |: | |    ||     :|  :|: ::
Db      95 SDRCTSRFGRRRPFIAAGAALVAVAVGLI---GFAADIGAASGDPTGNVAKPRAIAVFVV 151

Qy     128 GVGLLDFCGQVCFTPLEALLSDLFR-DPDHCRQAYSVYAFMISLGGCLGYLLPAID---- 182
           |  :||       |  |||:|:       | | : ::| ::||   ||   :
Db     152 GFWILDVANNTLQGPCRALLADMAAGSQTKTRYANAFFSFFMALGNIGGYAAGSYSRLYT 211

Qy     183 ----WDTSALAPYLGTQEECLFGLLT-LIFLTCVAATLLVAEEAALGPTEPAEGLSAPSL 237
               |:|   |    : | |  :| || || :| :::   :  :   :  | |   :
Db     212 VFPFTKTAACDVYCANLKSCFFISITLLIVLTILALSVVKERQITIDEIQEEEDLKNRNN 271

Qy     238 SPHCCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVG 297
           |  |    |||  |   | |:  |  |    |: :  | :     :|:|   | || ||::|
Db     272 SSGC----ARLPF--FGQLIGALKDL----PKPMLILLLVTALNWIAWFPFLLFDTDWMG 321

Qy     298 EGLYQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFG--TRAVY 355
           : :| |           : ||:||  |:||| :   :  | || :: | :  |   |
Db     322 KEVYGGT------VGEGKLYDQGVHAGALGLMINSVVLGVMSLSIEGLARMVGGAKRLWG 375

Qy     356 LASVAAFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYR 415
           : ::        | ||: :| :|| ||                    : |:    :
Db     376 IVNIIL------AVCLAMTV-LVTKSAE------------------HFRDSHHIM---- 405

Qy     416 GDTGGASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVG 475
```

```
                                       || : |||||  |     :::  |:|
Db    406 ---------------------------------GSAVPPPPPA--GVKGGALAIFAVLG 429

Qy    476 EPTEARV-VP----------GRGICLDLAILDSAFLLSQV----------------AP 506
          |      :|          | |  | | :|: | :: |:                  |
Db    430 IPLAITFSIPLALASIFSASSGSGQGLSLGVLNLAIVVPQMFVSVTSGPWDAMFGGGNLP 489

Qy    507 SLFMGSIVQLSQSVTAY 523
          :  :|::   : :| ::
Db    490 AFVVGAVAATASAVLSF 506


RESULT 11
US-09-439-313-547
; Sequence 547, Application US/09439313
; Patent No. 6329505
; GENERAL INFORMATION:
;  APPLICANT: Xu, Jiangchun
;  APPLICANT:  Dillon, Davin C.
;  APPLICANT:  Mitcham, Jennifer L.
;  APPLICANT:  Harlocker, Susan Louise
;  APPLICANT:  Jiang Yuqui
;  APPLICANT:  Reed, Steven G.
;  APPLICANT:  Kalos, Michael
;  APPLICANT:  Fanger, Gary
;  APPLICANT:  Retter, Mark
;  APPLICANT:  Solk, John
;  APPLICANT:  Day, Craig
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
;  TITLE OF INVENTION:  DIAGNOSIS OF PROSTATE CANCER
;  FILE REFERENCE: 210121.427C9
;  CURRENT APPLICATION NUMBER: US/09/439,313
;  CURRENT FILING DATE:  1999-11-12
;  NUMBER OF SEQ ID NOS: 575
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 547
;   LENGTH: 58
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-439-313-547


  Query Match             10.9%;  Score 312;  DB 4;  Length 58;
  Best Local Similarity   100.0%;  Pred. No. 1.8e-23;
  Matches   58;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    217 VAEEAALGPTEPAEGLSAPSLSPHCCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRL 274
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 VAEEAALGPTEPAEGLSAPSLSPHCCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRL 58


RESULT 12
US-09-439-313-564
; Sequence 564, Application US/09439313
; Patent No. 6329505
; GENERAL INFORMATION:
;  APPLICANT: Xu, Jiangchun
;  APPLICANT:  Dillon, Davin C.
;  APPLICANT:  Mitcham, Jennifer L.
;  APPLICANT:  Harlocker, Susan Louise
;  APPLICANT:  Jiang Yuqui
;  APPLICANT:  Reed, Steven G.
;  APPLICANT:  Kalos, Michael
;  APPLICANT:  Fanger, Gary
;  APPLICANT:  Retter, Mark
;  APPLICANT:  Solk, John
;  APPLICANT:  Day, Craig
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
;  TITLE OF INVENTION:  DIAGNOSIS OF PROSTATE CANCER
;  FILE REFERENCE: 210121.427C9
;  CURRENT APPLICATION NUMBER: US/09/439,313
;  CURRENT FILING DATE:  1999-11-12
;  NUMBER OF SEQ ID NOS: 575
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 564
;   LENGTH: 56
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-439-313-564


  Query Match             10.6%;  Score 304;  DB 4;  Length 56;
  Best Local Similarity   100.0%;  Pred. No. 1.1e-22;
  Matches   56;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    218 AEEAALGPTEPAEGLSAPSLSPHCCPCRARLAFRNLGALLPRLHQLCCRMPRTLRR 273
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 AEEAALGPTEPAEGLSAPSLSPHCCPCRARLAFRNLGALLPRLHQLCCRMPRTLRR 56


RESULT 13
US-09-071-710-37
; Sequence 37, Application US/09071710
; Patent No. 6130043
;  GENERAL INFORMATION:
;    APPLICANT:  BILLING-MEDEL, PATRICIA
;    APPLICANT:  COHEN, MAURICE
;    APPLICANT:  COLPITTS, TRACEY L.
;    APPLICANT:  FRIEDMAN, PAULA N.
;    APPLICANT:  GORDON, JULIAN
;    APPLICANT:  GRANADOS, EDWARD N.
;    APPLICANT:  HODGES, STEVEN C.
;    APPLICANT:  KLASS, MICHAEL R.
;    APPLICANT:  KRATOCHVIL, JON D.
;    APPLICANT:  ROBERTS-RAPP, LISA
;    APPLICANT:  RUSSELL, JOHN C.
;    APPLICANT:  STROUPE, STEPHEN D.
;    TITLE OF INVENTION:  REAGENTS AND METHODS USEFUL
;    TITLE OF INVENTION:  FOR DETECTING DISEASES OF THE PROSTATE
;    NUMBER OF SEQUENCES:  41
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Abbott Laboratories
;      STREET:  100 Abbott Park Road
;      CITY:  Abbott Park
;      STATE:  IL
;      COUNTRY:  USA
;      ZIP:  60064-3500
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/071,710
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/850,713
;      FILING DATE:  02-MAY-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Becker, Cheryl L.
;      REGISTRATION NUMBER:  35,441
;      REFERENCE/DOCKET NUMBER:  6083.US.P1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  847/935-1729
;      TELEFAX:  847/938-2623
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  37:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  44 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  No. 6130043e
US-09-071-710-37
```

Query Match 8.5%; Score 243; DB 4; Length 44;
Best Local Similarity 100.0%; Pred. No. 6.8e-17;
Matches 44; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 403 YHREKQVFLPKYRGDTGGASSEDSLMTSFLPGPKPGAPFPNGHV 446
||||||||||||||||||||||||||||||||||||||||||||
Db 1 YHREKQVFLPKYRGDTGGASSEDSLMTSFLPGPKPGAPFPNGHV 44

RESULT 14
US-09-525-397-37
; Sequence 37, Application US/09525397
; Patent No. 6252047
; GENERAL INFORMATION:
; APPLICANT: BILLING-MEDEL, PATRICIA
; APPLICANT: COHEN, MAURICE
; APPLICANT: COLPITTS, TRACEY L.
; APPLICANT: FRIEDMAN, PAULA N.
; APPLICANT: GORDON, JULIAN
; APPLICANT: GRANADOS, EDWARD N.
; APPLICANT: HODGES, STEVEN C.
; APPLICANT: KLASS, MICHAEL R.
; APPLICANT: KRATOCHVIL, JON D.
; APPLICANT: ROBERTS-RAPP, LISA
; APPLICANT: RUSSELL, JOHN C.
; APPLICANT: STROUPE, STEPHEN D.
; TITLE OF INVENTION: REAGENTS AND METHODS USEFUL
; TITLE OF INVENTION: FOR DETECTING DISEASES OF THE PROSTATE
; NUMBER OF SEQUENCES: 41
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Abbott Laboratories
; STREET: 100 Abbott Park Road
; CITY: Abbott Park
; STATE: IL
; COUNTRY: USA
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/525,397
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/071,710
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Becker, Cheryl L.
; REGISTRATION NUMBER: 35,441
; REFERENCE/DOCKET NUMBER: 6083.US.P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 847/935-1729
; TELEFAX: 847/938-2623
; TELEX:
; INFORMATION FOR SEQ ID NO: 37:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 44 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: No. 6252047e
US-09-525-397-37

Query Match 8.5%; Score 243; DB 4; Length 44;
Best Local Similarity 100.0%; Pred. No. 6.8e-17;
Matches 44; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 403 YHREKQVFLPKYRGDTGGASSEDSLMTSFLPGPKPGAPFPNGHV 446
||||||||||||||||||||||||||||||||||||||||||||

Db 1 YHREKQVFLPKYRGDTGGASSEDSLMTSFLPGPKPGAPFPNGHV 44

RESULT 15
US-09-439-313-546
; Sequence 546, Application US/09439313
; Patent No. 6329505
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan Louise
; APPLICANT: Jiang Yuqui
; APPLICANT: Reed, Steven G.
; APPLICANT: Kalos, Michael
; APPLICANT: Fanger, Gary
; APPLICANT: Retter, Mark
; APPLICANT: Solk, John
; APPLICANT: Day, Craig
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
; FILE REFERENCE: 210121.427C9
; CURRENT APPLICATION NUMBER: US/09/439,313
; CURRENT FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 575
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 546
; LENGTH: 29
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-439-313-546

Query Match 5.5%; Score 157; DB 4; Length 29;
Best Local Similarity 100.0%; Pred. No. 9.5e-09;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 295 FVGEGLYQGVPRAEPGTEARRHYDEGVRM 323
|||||||||||||||||||||||||||||
Db 1 FVGEGLYQGVPRAEPGTEARRHYDEGVRM 29

Search completed: June 26, 2002, 13:44:21
Job time: 4074 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: June 26, 2002, 12:34:42 ; Search time 80.68 Seconds
(without alignments)
761.326 Million cell updates/sec

Title: US-09-838-785-2
Perfect score: 2861
Sequence: 1 MVQRLWVSRLLRHRKAQLLL.....AIYFATQVVFDKSDLAKYSA 553

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_032802:*
1: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2861	100.0	553	19	AAW71869	Amino acid encoded
2	2861	100.0	553	19	AAW69385	Prostate tumour sp
3	2861	100.0	553	21	AAB28527	Protein encoded by
4	2861	100.0	553	21	AAY82002	Human immunogenic
5	2861	100.0	553	22	AAU69763	Human prostate cDN
6	2861	100.0	553	22	AAU04961	Human prostate tum
7	2861	100.0	553	22	AAM01117	Human prostate-spe
8	2861	100.0	553	22	AAG99002	Human prostate-spe
9	2861	100.0	553	22	AAG62150	Human P501S invent
10	2861	100.0	553	22	AAB74800	Prostate tumour an
11	2861	100.0	1079	22	AAB74830	Prostate tumour an
12	2596	90.7	595	22	AAM01318	Alpha prepro-P501S
13	1417.5	49.5	371	22	AAU69875	Human prostate cDN
14	1417.5	49.5	371	22	AAM01230	P553S splice varia
15	1416	49.5	371	22	AAE01362	Human gene 11 enco
16	1403.5	49.1	400	22	AAU69907	Human prostate pro
17	1403.5	49.1	400	22	AAM01262	Ra12-P501S-E2 cons
18	1287	45.0	255	20	AAW85068	Protein encoded by
19	1287	45.0	255	21	AAB29268	Human prostate-rel
20	1287	45.0	255	22	AAU04205	Prostate-specific
21	1150	40.2	231	21	AAY54369	Amino acid sequenc
22	1120	39.1	252	22	AAE01423	Human secreted pro
23	519.5	18.2	599	22	ABB60709	Drosophila melanog
24	457.5	16.0	748	22	AAM40227	Human polypeptide
25	426	14.9	123	22	AAU69873	Human prostate cDN
26	426	14.9	123	22	AAM01228	P553S splice varia
27	357	12.5	494	20	AAY41119	Soybean sucrose tr
28	337.5	11.8	523	20	AAY41129	Vicia faba sucrose
29	330	11.5	515	22	AAG80016	S. tuberosum SUT1
30	328.5	11.5	594	22	AAG80012	A. thaliana SUT2 p
31	326.5	11.4	667	20	AAY41118	Rice sucrose trans
32	325	11.4	563	20	AAY41125	Wheat sucrose tran
33	323.5	11.3	507	21	AAG30206	Arabidopsis thalia
34	323.5	11.3	512	21	AAG30205	Arabidopsis thalia
35	323.5	11.3	539	21	AAG30204	Arabidopsis thalia
36	313.5	11.0	605	22	AAG80018	S. tuberosum SUT2
37	313	10.9	497	20	AAY41116	Corn sucrose trans
38	312	10.9	58	22	AAU69833	Human prostate pro
39	312	10.9	58	22	AAM01188	Human prostate-spe
40	312	10.9	58	22	AAG99073	Human prostate-spe
41	311	10.9	533	20	AAY41128	Ricinus communis s
42	307.5	10.7	500	22	AAG80011	L. esculentum SUT4
43	307.5	10.7	500	22	AAG80017	S. tuberosum SUT4
44	305.5	10.7	501	20	AAY41126	Daucus carota sucr
45	303.5	10.6	519	20	AAY41114	Corn sucrose trans

ALIGNMENTS

RESULT 1
AAW71869
ID AAW71869 standard; Protein; 553 AA.
XX
AC AAW71869;
XX
DT 06-JAN-1999 (first entry)
XX
DE Amino acid encoded by prostate tumour clone L1-12.
XX
KW Prostate; cancer; tumour; vaccine; immunogen; clone.
XX
OS Homo sapiens.
XX
PN WO9837093-A2.
XX
PD 27-AUG-1998.
XX
PF 25-FEB-1998; 98WO-US03492.
XX
PR 09-FEB-1998; 98US-0020956.
PR 25-FEB-1997; 97US-0806099.
PR 01-AUG-1997; 97US-0904804.
XX
PA (CORI-) CORIXA CORP.
XX
PI Dillon DC, Xu J;
XX
DR WPI; 1998-609886/51.
DR N-PSDB; AAV61201.
XX
PT Polypeptides comprising immunogenic portions of prostate proteins -
PT used in a vaccine for the treatment of prostate cancer
XX

```
PS   Example 1; Page 82-84; 130pp; English.
XX
CC   The present sequence is an immunogenic portion of a prostate tumour
CC   protein. The immunogen, or the DNA encoding it, can be used as a
CC   vaccine for the treatment of prostate cancer. The immunogen was
CC   isolated from a prostate tumour cDNA library obtained by subtracting
CC   a prostate tumour cDNA expression library with a normal tissue cDNA
CC   library.
XX
SQ   Sequence   553 AA;
```

```
  Query Match            100.0%;  Score 2861;  DB 19;  Length 553;
  Best Local Similarity  100.0%;  Pred. No. 1.8e-269;
  Matches  553;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 MVQRLWVSRLLRHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 mvqrlwvsrllrhrkaqlllvnlltfglevclaagityvppllievgveekfmtmvlgig 60

Qy     61 PVLGLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDPRPL 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 pvlglvcvpllgsasdhwrgrygrrrpfiwalslgillslflipragwlagllcpdprpl 120

Qy    121 ELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPA 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 elallilgvglldfcgqvcftpleallsdlfrdpdhcrqaysvyafmislggclgyllpa 180

Qy    181 IDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPH 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 idwdtsalapylgtqeeclfglltlifltcvaatllvaeeaalgptepaeglsapslsph 240

Qy    241 CCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGL 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 ccpcrarlafrnlgallprlhqlccrmprtlrrlfvaelcswmalmtftlfytdfvgegl 300

Qy    301 YQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVA 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 yqgvpraepgtearrhydegvrmgslglflqcaislvfslvmdrlvqrfgtravylasva 360

Qy    361 AFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 afpvaagatclshsvavvtasaaltgftfsalqilpytlaslyhrekqvflpkyrgdtgg 420

Qy    421 ASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 assedslmtsflpgpkpgapfpnghvgaggsgllppppalcgasacdvsvrvvvgeptea 480

Qy    481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFATQ 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 rvvpgrgicldlaildsafllsqvapslfmgsivqlsqsvtaymvsaaglglvaiyfatq 540

Qy    541 VVFDKSDLAKYSA 553
          |||||||||||||
Db    541 vvfdksdlakysa 553
```

```
RESULT   2
AAW69385
ID   AAW69385 standard; Protein; 553 AA.
XX
AC   AAW69385;
XX
DT   08-DEC-1998  (first entry)
XX
DE   Prostate tumour specific gene clone L1-12 protein.
XX
KW   Prostate tumour specific gene; human; prostate cancer; detection;
KW   therapy.
XX
OS   Homo sapiens.
XX
PN   WO9837418-A2.
XX
PD   27-AUG-1998.
XX
PF   25-FEB-1998;   98WO-US03690.
XX
PR   09-FEB-1998;   98US-0904809.
PR   25-FEB-1997;   97US-0806596.
PR   01-AUG-1997;   97US-0904809.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Dillon DC,  Xu J;
XX
DR   WPI; 1998-480805/41.
DR   N-PSDB; AAV58586.
XX
PT   Novel human prostate specific tumour protein and fragments - useful
PT   for detecting and treating prostate cancers
XX
PS   Example 1; Page 87-89; 141pp; English.
XX
CC   This sequence is encoded by a human prostate tumour specific gene, and
CC   can be used in the method of the invention. The method is for detecting
CC   prostate cancer comprises contacting a biological sample with an agent
CC   able to bind an immunogenic portion of a prostate protein (such as
CC   this protein sequence). An antibody which binds to an immunogenic
CC   portion of the prostate protein, and the method can be used to detect,
CC   monitor progression of, or treat prostate cancers. The antibody may
CC   also be conjugated to a therapeutic agent for use in therapy of prostate
CC   cancers.
XX
SQ   Sequence   553 AA;
```

```
  Query Match            100.0%;  Score 2861;  DB 19;  Length 553;
  Best Local Similarity  100.0%;  Pred. No. 1.8e-269;
  Matches  553;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 MVQRLWVSRLLRHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 mvqrlwvsrllrhrkaqlllvnlltfglevclaagityvppllievgveekfmtmvlgig 60

Qy     61 PVLGLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDPRPL 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 pvlglvcvpllgsasdhwrgrygrrrpfiwalslgillslflipragwlagllcpdprpl 120

Qy    121 ELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPA 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 elallilgvglldfcgqvcftpleallsdlfrdpdhcrqaysvyafmislggclgyllpa 180

Qy    181 IDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPH 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 idwdtsalapylgtqeeclfglltlifltcvaatllvaeeaalgptepaeglsapslsph 240

Qy    241 CCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGL 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 ccpcrarlafrnlgallprlhqlccrmprtlrrlfvaelcswmalmtftlfytdfvgegl 300

Qy    301 YQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVA 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 yqgvpraepgtearrhydegvrmgslglflqcaislvfslvmdrlvqrfgtravylasva 360

Qy    361 AFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 afpvaagatclshsvavvtasaaltgftfsalqilpytlaslyhrekqvflpkyrgdtgg 420

Qy    421 ASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 assedslmtsflpgpkpgapfpnghvgaggsgllppppalcgasacdvsvrvvvgeptea 480

Qy    481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFATQ 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 rvvpgrgicldlaildsafllsqvapslfmgsivqlsqsvtaymvsaaglglvaiyfatq 540
```

```
Qy    541 VVFDKSDLAKYSA 553
          |||||||||||||
Db    541 vvfdksdlakysa 553


RESULT   3
AAB28527
ID   AAB28527 standard; Protein; 553 AA.
XX
AC   AAB28527;
XX
DT   07-FEB-2001  (first entry)
XX
DE   Protein encoded by human breast tumour cDNA clone P501S.
XX
KW   Human; breast tumour antigen; cytostatic; immunotherapy;
KW   breast cancer; vaccine.
XX
OS   Homo sapiens.
XX
PN   WO200061756-A2.
XX
PD   19-OCT-2000.
XX
PF   10-APR-2000; 2000WO-US09688.
XX
PR   09-APR-1999;   99US-0288950.
PR   02-JUL-1999;   99US-0346327.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Reed SG,  Xu J,  Dillon DC;
XX
DR   WPI; 2000-638568/61.
DR   N-PSDB; AAC79473.
XX
PT   A novel isolated polypeptide comprising an immunogenic portion of a
PT   breast cancer protein useful in the detection and treatment of breast
PT   cancer -
XX
PS   Claim 2; Page 92-93; 95pp; English.
XX
CC   The present sequence is encoded by a cDNA sequence which was isolated
CC   from a breast tumour cDNA library. It is provided in a specification .
CC   relating to compounds for immunotherapy and diagnosis of breast cancer.
CC   Breast tumour antigens and the polynucleotides that encode them may be
CC   used in the production of a pharmaceutical composition to be used in the
CC   treatment of breast cancer. Proliferated T cells and incubated antigen
CC   presenting cells are also required. The polypeptides and polynucleotides
CC   may also be used to produce a vaccine.
XX
SQ   Sequence   553 AA;


  Query Match             100.0%;  Score 2861;  DB 21;  Length 553;
  Best Local Similarity   100.0%;  Pred. No. 1.8e-269;
  Matches  553;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 MVQRLWVSRLLRHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIG 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 mvqrlwvsrllrhrkaqlllvnlltfglevclaagityvpplllevgveekfmtmvlgig 60

Qy     61 PVLGLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDPRPL 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 pvlglvcvpllgsasdhwrgrygrrrpfiwalslgillslflipragwlagllcpdprpl 120

Qy    121 ELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPA 180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 elallilgvglldfcgqvcftpleallsdlfrdpdhcrqaysvyafmislggclgyllpa 180

Qy    181 IDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPH 240
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 idwdtsalapylgtqeeclfglltlifltcvaatllvaeeaalgptepaeglsapslsph 240

Qy    241 CCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGL 300
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 ccpcrarlafrnlgallprlhqlccrmprtlrrlfvaelcswmalmtftlfytdfvgegl 300

Qy    301 YQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVA 360
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 yqgvpraepgtearrhydegvrmgslglflqcaislvfslvmdrlvqrfgtravylasva 360

Qy    361 AFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 afpvaagatclshsvavvtasaaltgftfsalqilpytlaslyhrekqvflpkyrgdtgg 420

Qy    421 ASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 assedslmtsflpgpkpgapfpnghvgaggsgllppppalcgasacdvsvrvvvgeptea 480

Qy    481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFATQ 540
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 rvvpgrgicldlaildsafllsqvapslfmgsivqlsqsvtaymvsaaglglvaiyfatq 540

Qy    541 VVFDKSDLAKYSA 553
          |||||||||||||
Db    541 vvfdksdlakysa 553


RESULT   4
AAY82002
ID   AAY82002 standard; Protein; 553 AA.
XX
AC   AAY82002;
XX
DT   13-JUN-2000  (first entry)
XX
DE   Human immunogenic prostate tumour protein sequence SEQ ID NO:113.
XX
KW   Human; prostate cancer; diagnosis; tumour; gene therapy; detection;
KW   immunogenic; cytostatic; vaccine.
XX
OS   Homo sapiens.
XX
PN   WO200004149-A2.
XX
PD   27-JAN-2000.
XX
PF   14-JUL-1999;   99WO-US15838.
XX
PR   14-JUL-1998;   98US-0115453.
PR   14-JUL-1998;   98US-0116134.
PR   23-SEP-1998;   98US-0159812.
PR   23-SEP-1998;   98US-0159822.
PR   15-JAN-1999;   99US-0232149.
PR   15-JAN-1999;   99US-0232880.
PR   09-APR-1999;   99US-0288946.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Dillon DC,  Harlocker SL,  Yuqiu J,  Xu J,  Mitcham JL;
XX
DR   WPI; 2000-171268/15.
XX
PT   New polypeptide useful for treating and diagnosing prostate cancer
PT   comprises an immunogenic portion of prostate tumor protein -
XX
PS   Claim 3; Page 138-139; 263pp; English.
XX
CC   The present invention describes isolated polypeptides, comprising an
CC   immunogenic portion of a prostate tumour protein (PTP). The polypeptides
CC   and polynucleotides encoding them have cytostatic activity and can be
CC   used in vaccines and in gene therapy. The polypeptides and
CC   polynucleotides encoding them, antigen presenting cells which express
CC   the polypeptides, antibodies against the polypeptides and vaccines
```

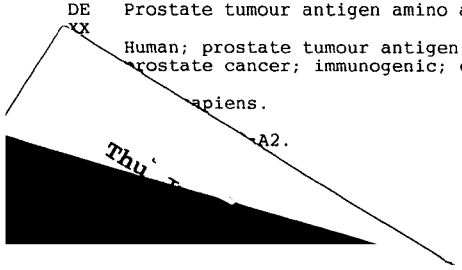
```
CC   comprising them can be used for inhibiting the development of prostate
CC   cancer in a patient. The polypeptides can be used to generate antibodies
CC   or anti-idiotypic antibodies for passive immuno therapy. A portion of
CC   the polynucleotides encoding the polypeptides can be used as a probe or
CC   to modulate the expression of the polypeptides. AAA06241 to AAA06691 and
CC   AAY82000 to AAY82020 represent sequences used in the exemplification of
CC   the present invention.
XX
SQ   Sequence   553 AA;


  Query Match             100.0%;  Score 2861;  DB 21;  Length 553;
  Best Local Similarity   100.0%;  Pred. No. 1.8e-269;
  Matches  553;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 MVQRLWVSRLLRHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 mvqrlwvsrllrhrkaqlllvnlltfglevclaagityvppllevgveekfmtmvlgig 60

Qy      61 PVLGLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDPRPL 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 pvlglvcvpllgsasdhwrgrygrrrpfiwalslgillslflipragwlagllcpdprpl 120

Qy     121 ELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPA 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 elallilgvglldfcgqvcftpleallsdlfrdpdhcrqaysvyafmislggclgyllpa 180

Qy     181 IDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPH 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 idwdtsalapylgtqeeclfglltlifltcvaatllvaeeaalgptepaeglsapslsph 240

Qy     241 CCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGL 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 ccpcrarlafrnlgallprlhqlccrmprtlrrlfvaelcswmalmtftlfytdfvgegl 300

Qy     301 YQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVA 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 yqgvpraepgtearrhydegvrmgslglflqcaislvfslvmdrlvqrfgtravylasva 360

Qy     361 AFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 afpvaagatclshsvavvtasaaltgftfsalqilpytlaslyhrekqvflpkyrgdtgg 420

Qy     421 ASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 assedslmtsflpgpkpgapfpnghvgaggsgllppppalcgasacdvsvrvvvgeptea 480

Qy     481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFATQ 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 rvvpgrgicldlaildsafllsqvapslfmgsivqlsqsvtaymvsaaglglvaiyfatq 540

Qy     541 VVFDKSDLAKYSA 553
          |||||||||||||
Db     541 vvfdksdlakysa 553


RESULT   5
AAU69763
ID   AAU69763 standard; Protein; 553 AA.
XX
AC   AAU69763;
XX
DT   30-JAN-2002  (first entry)
XX
DE   Human prostate cDNA encoded protein #3.
XX
KW   Human; prostate cancer; cytostatic; immunostimulant; tumour; immunogen.
XX
OS   Homo sapiens.
XX
PN   WO200173032-A2.
XX
PD   04-OCT-2001.
XX
PF   27-MAR-2001; 2001WO-US09919.
XX
PR   27-MAR-2000; 2000US-0536857.
PR   09-MAY-2000; 2000US-0568100.
PR   12-MAY-2000; 2000US-0570737.
PR   13-JUN-2000; 2000US-0593793.
PR   27-JUN-2000; 2000US-0605783.
PR   10-AUG-2000; 2000US-0636215.
PR   29-AUG-2000; 2000US-0651236.
PR   06-SEP-2000; 2000US-0657279.
PR   02-OCT-2000; 2000US-0679426.
PR   10-OCT-2000; 2000US-0685166.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Kalos MD;
PI   Fanger GR, Retter MW, Stolk JA, Day CH, Vedvick TS, Carter D;
PI   Li SX, Wang A, Skeiky YAW, Hepler WT, Henderson RA;
XX
DR   WPI; 2001-639232/73.
DR   N-PSDB; AAS63557.
XX
PT   New human prostate-specific polypeptides and polynucleotides useful for
PT   the diagnosis and treatment of cancer, especially prostate cancer -
XX
PS   Claim 2; Page 269-270; 579pp; English.
XX
CC   The invention relates to isolated prostate-specific
CC   polynucleotides, polypeptides, fusion proteins of the polypeptides,
CC   antibodies raised against the polypeptides (or antigenic epitopes
CC   derived from them) and antigen-presenting cells expressing the
CC   polypeptides. The antibodies are useful for detecting the presence of
CC   cancer, especially prostate cancer. The polypeptides, polynucleotides and
CC   the antigen-presenting cells are useful for stimulating and/or expanding
CC   T cells specific for a tumour protein, and for inhibiting the development
CC   of cancer especially prostate cancer. Compositions comprising the
CC   polynucleotide and/or polypeptide are useful for stimulating an immune
CC   response, and for treating cancer. The oligonucleotide is useful for
CC   detecting cancer. The present sequence is a prostate specific
CC   polypeptide of the invention.
XX
SQ   Sequence   553 AA;


  Query Match             100.0%;  Score 2861;  DB 22;  Length 553;
  Best Local Similarity   100.0%;  Pred. No. 1.8e-269;
  Matches  553;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 MVQRLWVSRLLRHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 mvqrlwvsrllrhrkaqlllvnlltfglevclaagityvppllevgveekfmtmvlgig 60

Qy      61 PVLGLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDPRPL 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 pvlglvcvpllgsasdhwrgrygrrrpfiwalslgillslflipragwlagllcpdprpl 120

Qy     121 ELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPA 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 elallilgvglldfcgqvcftpleallsdlfrdpdhcrqaysvyafmislggclgyllpa 180

Qy     181 IDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPH 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 idwdtsalapylgtqeeclfglltlifltcvaatllvaeeaalgptepaeglsapslsph 240

Qy     241 CCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGL 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 ccpcrarlafrnlgallprlhqlccrmprtlrrlfvaelcswmalmtftlfytdfvgegl 300

Qy     301 YQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVA 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 yqgvpraepgtearrhydegvrmgslglflqcaislvfslvmdrlvqrfgtravylasva 360
```

```
Qy    361 AFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 afpvaagatclshsvavvtasaaltgftfsalqilpytlaslyhrekqvflpkyrgdtgg 420

Qy    421 ASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 assedslmtsflpgpkpgapfpnghvgaggsgllppppalcgasacdvsvrvvvgeptea 480

Qy    481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFATQ 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 rvvpgrgicldlaildsafllsqvapslfmgsivqlsqsvtaymvsaaglglvaiyfatq 540

Qy    541 VVFDKSDLAKYSA 553
          |||||||||||||
Db    541 vvfdksdlakysa 553


RESULT   6
AAU04961
ID   AAU04961 standard; Protein; 553 AA.
XX
AC   AAU04961;
XX
DT   24-OCT-2001  (first entry)
XX
DE   Human prostate tumour protein L1-12.
XX
KW   Human; prostate tumour protein; prostate cancer.
XX
OS   Homo sapiens.
XX
PN   US6262245-B1.
XX
PD   17-JUL-2001.
XX
PF   25-FEB-1998;   98US-0030607.
XX
PR   25-FEB-1997;   97US-0806099.
PR   01-AUG-1997;   97US-0904804.
PR   09-FEB-1998;   98US-0020956.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Xu J,  Dillon DC;
XX
DR   WPI; 2001-440862/47.
DR   N-PSDB; AAS10108.
XX
PT   Novel polynucleotide encoding polypeptide comprising a portion of
PT   prostate tumour protein useful for inhibiting development of prostate
PT   cancer or for treating prostate cancer in a patient -
XX
PS   Example 1; Column 125-127; 105pp; English.
XX
CC   The sequence is a partial prostate tumour protein, encoded by a prostate
CC   tumour specific cDNA. The DNA is useful for inhibiting the development
CC   of prostate cancer or for treating prostate cancer in a patient.
XX
SQ   Sequence   553 AA;


  Query Match             100.0%;  Score 2861;  DB 22;  Length 553;
  Best Local Similarity   100.0%;  Pred. No. 1.8e-269;
  Matches  553;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 MVQRLWVSRLLRHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 mvqrlwvsrllrhrkaqlllvnlltfglevclaagityvppllievgveekfmtmvlgig 60

Qy     61 PVLGLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDPRPL 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 pvlglvcvpllgsasdhwrgrygrrrpfiwalslgillslflipragwlagllcpdprpl 120

Qy    121 ELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPA 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 elallilgvglldfcgqvcftpleallsdlfrdpdhcrqaysvyafmislggclgyllpa 180

Qy    181 IDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPH 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 idwdtsalapylgtqeeclfglltlifltcvaatllvaeeaalgptepaeglsapslsph 240

Qy    241 CCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGL 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 ccpcrarlafrnlgallprlhqlccrmprtlrrlfvaelcswmalmtftlfytdfvgegl 300

Qy    301 YQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVA 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 yqgvpraepgtearrhydegvrmgslglflqcaislvfslvmdrlvqrfgtravylasva 360

Qy    361 AFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 afpvaagatclshsvavvtasaaltgftfsalqilpytlaslyhrekqvflpkyrgdtgg 420

Qy    421 ASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 assedslmtsflpgpkpgapfpnghvgaggsgllppppalcgasacdvsvrvvvgeptea 480

Qy    481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFATQ 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 rvvpgrgicldlaildsafllsqvapslfmgsivqlsqsvtaymvsaaglglvaiyfatq 540

Qy    541 VVFDKSDLAKYSA 553
          |||||||||||||
Db    541 vvfdksdlakysa 553


RESULT   7
AAM01117
ID   AAM01117 standard; Protein; 553 AA.
XX
AC   AAM01117;
XX
DT   04-OCT-2001  (first entry)
XX
DE   Human prostate-specific amino acid sequence L1-12.
XX
KW   Human; prostate cancer; prostate-specific; diagnosis; vaccine;
KW   cytostatic; gene therapy; metastasis.
XX
OS   Homo sapiens.
XX
PN   WO200151633-A2.
XX
PD   19-JUL-2001.
XX
PF   16-JAN-2001; 2001WO-US01574.
XX
PR   14-JAN-2000; 2000US-0483672.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Xu J,  Dillon DC,  Mitcham JL,  Harlocker SL,  Jiang Y,  Reed SG;
PI   Kalos MD,  Fanger GR,  Day CH,  Retter MW,  Stolk JA,  Skeiky YAW;
PI   Wang A,  Meagher MJ;
XX
DR   WPI; 2001-425873/45.
XX
PT   New polynucleotide encoding a prostate-specific protein, for
PT   diagnosing, monitoring and treating prostate cancer in a patient and
PT   for use in vaccines -
XX
PS   Claim 2; Page 267-268; 543pp; English.
XX
CC   The present invention describes polynucleotide sequences (I) which encode
CC   prostate-specific proteins (II). (I) and (II) have cytostatic activity,
CC   and can be used in vaccine production and gene therapy. (I), (II),
```

```
CC   antibodies to (II), fusion proteins comprising (II), and isolated
CC   T cells prepared using (I) or (II) are used treat cancer in a patient.
CC   (I) and the antibodies are also used in the detection of cancer in a
CC   patient. The cancer that is diagnosed or treated is particularly
CC   prostate cancer. (I) and (II) can be used in vaccines. The antibodies or
CC   (I) can be used for monitoring the progression of cancer in a patient.
CC   (I) and (II) can also be used to improve diagnostic and therapeutic
CC   methods for prostate cancer. They can indicate the level of metastasis
CC   as well as the prostate volume. AAH93357 to AAH93944 and AAM01115 to
CC   AAM01318 represent polynucleotide and amino acid sequences used in the
CC   exemplification of the present invention.
XX
SQ   Sequence   553 AA;


  Query Match           100.0%;  Score 2861;  DB 22;  Length 553;
  Best Local Similarity  100.0%;  Pred. No. 1.8e-269;
  Matches  553;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 MVQRLWVSRLLRHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 mvqrlwvsrllrhrkaqlllvnlltfglevclaagityvppllevgveekfmtmvlgig 60

Qy     61 PVLGLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDPRPL 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 pvlglvcvpllgsasdhwrgrygrrrpfiwalslgillslflipragwlagllcpdprpl 120

Qy    121 ELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPA 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 elallilgvglldfcgqvcftpleallsdlfrdpdhcrqaysvyafmislggclgyllpa 180

Qy    181 IDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPH 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 idwdtsalapylgtqeeclfglltlifltcvaatllvaeeaalgptepaeglsapslsph 240

Qy    241 CCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGL 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 ccpcrarlafrnlgallprlhqlccrmprtlrrlfvaelcswmalmtftlfytdfvgegl 300

Qy    301 YQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVA 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 yqgvpraepgtearrhydegvrmgslglflqcaislvfslvmdrlvqrfgtravylasva 360

Qy    361 AFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 afpvaagatclshsvavvtasaaltgftfsalqilpytlaslyhrekqvflpkyrgdtgg 420

Qy    421 ASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 assedslmtsflpgpkpgapfpnghvgaggsgllppppalcgasacdvsvrvvvgeptea 480

Qy    481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFATQ 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 rvvpgrgicldlaildsafllsqvapslfmgsivqlsqsvtaymvsaaglglvaiyfatq 540

Qy    541 VVFDKSDLAKYSA 553
          |||||||||||||
Db    541 vvfdksdlakysa 553


RESULT   8
AAG99002
ID   AAG99002 standard; Protein; 553 AA.
XX
AC   AAG99002;

XX
DT   25-SEP-2001  (first entry)

DE   Human prostate-specific amino acid sequence L1-12/P501S.

KW   Human; prostate cancer; therapy; diagnosis; cat eye syndrome;
KW   chromosome 22q11.2; prostate-specific protein; chromosome 1;
KW   prostatic antigen; PSA.

XX
OS   Homo sapiens.
XX
PN   WO200134802-A2.
XX
PD   17-MAY-2001.
XX
PF   09-NOV-2000; 2000WO-US30904.
XX
PR   12-NOV-1999;   99US-0439313.
PR   18-NOV-1999;   99US-0443686.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Xu J,  Dillon DC,  Mitcham JL,  Harlocker SL,  Jiang Y,  Reed SG;
PI   Kalos MD,  Retter MW,  Stolk JA,  Day CH,  Skeiky YAW,  Wang A;
XX
DR   WPI; 2001-308785/32.
XX
PT   Isolated polypeptide comprising at least an immunogenic portion of a
PT   prostate-specific protein, useful in the diagnosis and therapy of
PT   prostate cancer -
XX
PS   Claim 3; Page 167-168; 325pp; English.
XX
CC   The present invention describes an isolated polypeptide (P1) comprising
CC   at least an immunogenic portion of a prostate-specific protein, or its
CC   variant. Also described are polynucleotides (N1) encoding (P1). (P1) and
CC   (N1) have cytostatic activity and can be used in vaccine production.
CC   The polypeptides, nucleic acids and antibodies from the present
CC   invention are useful in the diagnosis and therapy of prostate cancer.
CC   Prostate specific genes P704P, P712P, P774P, P775P and B305D are located
CC   in a genomic region on chromosome 22q11.2 known as the Cat Eye Syndrome
CC   region. Prostate specific antigen (PSA) P501S was located on
CC   chromosome 1. AAH84671 to AAH85143 and AAG99000 to AAG99077 represent
CC   polynucleotide and polypeptide sequences used in the exemplification
CC   of the present invention.
XX
SQ   Sequence   553 AA;


  Query Match           100.0%;  Score 2861;  DB 22;  Length 553;
  Best Local Similarity  100.0%;  Pred. No. 1.8e-269;
  Matches  553;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 MVQRLWVSRLLRHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 mvqrlwvsrllrhrkaqlllvnlltfglevclaagityvppllevgveekfmtmvlgig 60

Qy     61 PVLGLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDPRPL 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 pvlglvcvpllgsasdhwrgrygrrrpfiwalslgillslflipragwlagllcpdprpl 120

Qy    121 ELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPA 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 elallilgvglldfcgqvcftpleallsdlfrdpdhcrqaysvyafmislggclgyllpa 180

Qy    181 IDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPH 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 idwdtsalapylgtqeeclfglltlifltcvaatllvaeeaalgptepaeglsapslsph 240

Qy    241 CCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGL 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 ccpcrarlafrnlgallprlhqlccrmprtlrrlfvaelcswmalmtftlfytdfvgegl 300

Qy    301 YQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVA 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 yqgvpraepgtearrhydegvrmgslglflqcaislvfslvmdrlvqrfgtravylasva 360

Qy    361 AFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 afpvaagatclshsvavvtasaaltgftfsalqilpytlaslyhrekqvflpkyrgdtgg 420
```

```
Qy    421 ASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 assedslmtsflpgpkpgapfpnghvgaggsgllppppalcgasacdvsvrvvvgeptea 480

Qy    481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFATQ 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 rvvpgrgicldlaildsafllsqvapslfmgsivqlsqsvtaymvsaaglglvaiyfatq 540

Qy    541 VVFDKSDLAKYSA 553
          |||||||||||||
Db    541 vvfdksdlakysa 553


RESULT   9
AAG62150
ID   AAG62150 standard; Protein; 553 AA.
XX
AC   AAG62150;
XX
DT   06-JUL-2001  (first entry)
XX
DE   Human P501S inventive antigen SEQ ID NO: 333.
XX
KW   Human; mouse; immunotherapy; cancer; leukaemia; WT1; Wilm's tumour gene;
KW   chromosome 11p13; zinc finger transcription factor.
XX
OS   Homo sapiens.
XX
PN   WO200125273-A2.
XX
PD   12-APR-2001.
XX
PF   04-OCT-2000; 2000WO-US27465.
XX
PR   04-OCT-1999;   99US-0157459.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Skeiky YAW,  Xu J,  Cheever MA,  Reed SG;
XX
DR   WPI; 2001-328324/34.
XX
PT   Polypeptide comprising part of the Wilm's Tumour gene product sequence is
PT   used in the diagnosis and treatment of malignant diseases e.g. leukemia
PT   and cancer associated with WTI  -
XX
PS   Disclosure; Page 212-213; 228pp; English.
XX
CC   The present invention describes compositions comprising peptides derived
CC   from the Wilm's tumour protein WT1 and methods for their use in treating
CC   malignant diseases. Peptides derived from both the murine and human WT1
CC   proteins are provided. The human WT1 gene is found on chromosome 11p13,
CC   and the protein was shown to be a zinc finger transcription factor. The
CC   immunogenic peptides of the invention are particularly useful in the
CC   diagnosis and treatment of cancer and leukaemia. The present sequence is
CC   a polypeptide described in the exemplification of the invention.
XX
SQ   Sequence   553 AA;


  Query Match            100.0%;  Score 2861;  DB 22;  Length 553;
  Best Local Similarity  100.0%;  Pred. No. 1.8e-269;
  Matches  553;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 MVQRLWVSRLLRHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 mvqrlwvsrllrhrkaqlllvnlltfglevclaagityvpplllevgveekfmtmvlgig 60

Qy     61 PVLGLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDPRPL 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 pvlglvcvpllgsasdhwrgrygrrrpfiwalslgillslflipragwlagllcpdprpl 120

Qy    121 ELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPA 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 elallilgvglldfcgqvcftpleallsdlfrdpdhcrqaysvyafmislggclgyllpa 180

Qy    181 IDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPH 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 idwdtsalapylgtqeeclfglltlifltcvaatllvaeeaalgptepaeglsapslsph 240

Qy    241 CCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGL 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 ccpcrarlafrnlgallprlhqlccrmprtlrrlfvaelcswmalmtftlfytdfvgegl 300

Qy    301 YQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVA 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 yqgvpraepgtearrhydegvrmgslglflqcaislvfslvmdrlvqrfgtravylasva 360

Qy    361 AFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 afpvaagatclshsvavvtasaaltgftfsalqilpytlaslyhrekqvflpkyrgdtgg 420

Qy    421 ASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 assedslmtsflpgpkpgapfpnghvgaggsgllppppalcgasacdvsvrvvvgeptea 480

Qy    481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFATQ 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 rvvpgrgicldlaildsafllsqvapslfmgsivqlsqsvtaymvsaaglglvaiyfatq 540

Qy    541 VVFDKSDLAKYSA 553
          |||||||||||||
Db    541 vvfdksdlakysa 553


RESULT  10
AAB74800
ID   AAB74800 standard; Protein; 553 AA.
XX
AC   AAB74800;
XX
DT   14-JUN-2001  (first entry)
XX
DE   Prostate tumour antigen predicted amino acid sequence for L1-12.
XX
KW   Human; prostate tumour antigen; prostate tumour; therapy; diagnosis;
KW   prostate cancer; immunogenic; cytostatic; vaccine.
XX
OS   Homo sapiens.
XX
PN   WO200125272-A2.
XX
PD   12-APR-2001.
XX
PF   04-OCT-2000; 2000WO-US27464.
XX
PR   04-OCT-1999;   99US-0157455.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Xu J,  Skeiky YAW,  Reed SG,  Cheever MA;
XX
DR   WPI; 2001-245062/25.
DR   N-PSDB; AAH02530.
XX
PT   Prostate specific protein and its encoding polynucleotide, useful for
PT   the treatment and diagnosis of prostate cancer -
XX
PS   Claim 3; Page 157-158; 276pp; English.
XX
CC   The present invention describes an isolated polypeptide (I) comprising
CC   at least an immunogenic portion of a prostate tumour antigen protein or
CC   its variant. (I) have cytostatic activity and can be used in vaccine
CC   production. (I), prostate tumour antigen polynucleotides, an antigen
CC   presenting cell (APC e.g. a dendritic cell) that expresses (I), and a
CC   pharmaceutical composition containing (I) are useful for inhibiting the
```

```
CC   development of cancer in a patient. Antibodies specific for prostate
CC   specific proteins and oligonucleotides that hybridise to a
CC   polynucleotide that encodes a prostate specific protein are useful
CC   for detecting the presence or absence of a cancer or monitoring the
CC   progression the progression of a cancer, especially prostate cancer.
CC   AAH02422 to AAH2872, AAB74798 to AAB74821 and AAB74830 are sequences
CC   used in the exemplification of the present invention.
XX
SQ   Sequence   553 AA;


  Query Match           100.0%;  Score 2861;  DB 22;  Length 553;
  Best Local Similarity  100.0%;  Pred. No. 1.8e-269;
  Matches  553;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 MVQRLWVSRLLRHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIG 60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 mvqrlwvsrllrhrkaqlllvnlltfglevclaagityvppllevgveekfmtmvlgig 60

Qy      61 PVLGLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDPRPL 120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 pvlglvcvpllgsasdhwrgrygrrrpfiwalslgillslflipragwlagllcpdprpl 120

Qy     121 ELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPA 180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 elallilgvglldfcgqvcftpleallsdlfrdpdhcrqaysvyafmislggclgyllpa 180

Qy     181 IDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPH 240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 idwdtsalapylgtqeeclfglltlifltcvaatllvaeeaalgptepaeglsapslsph 240

Qy     241 CCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGL 300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 ccpcrarlafrnlgallprlhqlccrmprtlrrlfvaelcswmalmtftlfytdfvgegl 300

Qy     301 YQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVA 360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 yqgvpraepgtearrhydegvrmgslglflqcaislvfslvmdrlvqrfgtravylasva 360

Qy     361 AFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 afpvaagatclshsvavvtasaaltgftfsalqilpytlaslyhrekqvflpkyrgdtgg 420

Qy     421 ASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 assedslmtsflpgpkpgapfpnghvgaggsgllppppalcgasacdvsvrvvvgeptea 480

Qy     481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFATQ 540
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 rvvpgrgicldlaildsafllsqvapslfmgsivqlsqsvtaymvsaaglglvaiyfatq 540

Qy     541 VVFDKSDLAKYSA 553
             |||||||||||||
Db     541 vvfdksdlakysa 553


RESULT  11
AAB74830
ID   AAB74830 standard; Protein; 1079 AA.
XX
AC   AAB74830;
XX
DT   14-JUN-2001  (first entry)
XX
DE   Prostate tumour antigen amino acid sequence for a fusion protein.
XX
     Human; prostate tumour antigen; prostate tumour; therapy; diagnosis;
     prostate cancer; immunogenic; cytostatic; vaccine.

     apiens.

     A2.
```



```
PD   12-APR-2001.
XX
PF   04-OCT-2000; 2000WO-US27464.
XX
PR   04-OCT-1999;   99US-0157455.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Xu J,  Skeiky YAW,  Reed SG,  Cheever MA;
XX
DR   WPI; 2001-245062/25.
XX
PT   Prostate specific protein and its encoding polynucleotide, useful for
PT   the treatment and diagnosis of prostate cancer -
XX
PS   Disclosure; Page 272-276; 276pp; English.
XX
CC   The present invention describes an isolated polypeptide (I) comprising
CC   at least an immunogenic portion of a prostate tumour antigen protein or
CC   its variant. (I) have cytostatic activity and can be used in vaccine
CC   production. (I), prostate tumour antigen polynucleotides, an antigen
CC   presenting cell (APC e.g. a dendritic cell) that expresses (I), and a
CC   pharmaceutical composition containing (I) are useful for inhibiting the
CC   development of cancer in a patient. Antibodies specific for prostate
CC   specific proteins and oligonucleotides that hybridise to a
CC   polynucleotide that encodes a prostate specific protein are useful
CC   for detecting the presence or absence of a cancer or monitoring the
CC   progression the progression of a cancer, especially prostate cancer.
CC   AAH02422 to AAH2872, AAB74798 to AAB74821 and AAB74830 are sequences
CC   used in the exemplification of the present invention.
XX
SQ   Sequence   1079 AA;


  Query Match           100.0%;  Score 2861;  DB 22;  Length 1079;
  Best Local Similarity  100.0%;  Pred. No. 4.5e-269;
  Matches  553;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 MVQRLWVSRLLRHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIG 60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     527 mvqrlwvsrllrhrkaqlllvnlltfglevclaagityvppllevgveekfmtmvlgig 586

Qy      61 PVLGLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDPRPL 120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     587 pvlglvcvpllgsasdhwrgrygrrrpfiwalslgillslflipragwlagllcpdprpl 646

Qy     121 ELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPA 180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     647 elallilgvglldfcgqvcftpleallsdlfrdpdhcrqaysvyafmislggclgyllpa 706

Qy     181 IDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPH 240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     707 idwdtsalapylgtqeeclfglltlifltcvaatllvaeeaalgptepaeglsapslsph 766

Qy     241 CCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGL 300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     767 ccpcrarlafrnlgallprlhqlccrmprtlrrlfvaelcswmalmtftlfytdfvgegl 826

Qy     301 YQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVA 360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     827 yqgvpraepgtearrhydegvrmgslglflqcaislvfslvmdrlvqrfgtravylasva 886

Qy     361 AFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     887 afpvaagatclshsvavvtasaaltgftfsalqilpytlaslyhrekqvflpkyrgdtgg 946

Qy     421 ASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     947 assedslmtsflpgpkpgapfpnghvgaggsgllppppalcgasacdvsvrvvvgeptea 1006

Qy     481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFATQ 540
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1007 rvvpgrgicldlaildsafllsqvapslfmgsivqlsqsvtaymvsaaglglvaiyfatq 1066
```

```
Qy     541 VVFDKSDLAKYSA 553
           |||||||||||||
Db    1067 vvfdksdlakysa 1079


RESULT  12
AAM01318
ID   AAM01318 standard; Protein; 595 AA.
XX
AC   AAM01318;
XX
DT   04-OCT-2001  (first entry)
XX
DE   Alpha prepro-P501S recombinant protein amino acid sequence.
XX
KW   Human; prostate cancer; prostate-specific; diagnosis; vaccine;
KW   cytostatic; gene therapy; metastasis.
XX
OS   Homo sapiens.
XX
PN   WO200151633-A2.
XX
PD   19-JUL-2001.
XX
PF   16-JAN-2001; 2001WO-US01574.
XX
PR   14-JAN-2000; 2000US-0483672.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Xu J,  Dillon DC,  Mitcham JL,  Harlocker SL,  Jiang Y,  Reed SG;
PI   Kalos MD,  Fanger GR,  Day CH,  Retter MW,  Stolk JA,  Skeiky YAW;
PI   Wang A,  Meagher MJ;
XX
DR   WPI; 2001-425873/45.
XX
PT   New polynucleotide encoding a prostate-specific protein, for
PT   diagnosing, monitoring and treating prostate cancer in a patient and
PT   for use in vaccines -
XX
PS   Example 17; Page 541-543; 543pp; English.
XX
CC   The present invention describes polynucleotide sequences (I) which encode
CC   prostate-specific proteins (II). (I) and (II) have cytostatic activity,
CC   and can be used in vaccine production and gene therapy. (I), (II),
CC   antibodies to (II), fusion proteins comprising (II), and isolated
CC   T cells prepared using (I) or (II) are used treat cancer in a patient.
CC   (I) and the antibodies are also used in the detection of cancer in a
CC   patient. The cancer that is diagnosed or treated is particularly
CC   prostate cancer. (I) and (II) can be used in vaccines. The antibodies or
CC   (I) can be used for monitoring the progression of cancer in a patient.
CC   (I) and (II) can also be used to improve diagnostic and therapeutic
CC   methods for prostate cancer. They can indicate the level of metastasis
CC   as well as the prostate volume. AAH93357 to AAH93944 and AAM01115 to
CC   AAM01318 represent polynucleotide and amino acid sequences used in the
CC   exemplification of the present invention.
XX
SQ   Sequence   595 AA;


  Query Match             90.7%;  Score 2596;  DB 22;  Length 595;
  Best Local Similarity   95.1%;  Pred. No. 1.2e-243;
  Matches  501;  Conservative    9;  Mismatches    9;  Indels    8;  Gaps    1;

Qy      35 GITYVPPLLLEVGVEEKFMT--------MVLGIGPVLGLVCVPLLGSASDHWRGRYGRRR 86
           |: ::   :  :  :|: ::         ||||||||||||||||||||||||||||||
Db      61 gllfinttiasiaakeegvslekreaeamvlgigpvlglvcvpllgsasdhwrgrygrrr 120

Qy      87 PFIWALSLGILLSLFLIPRAGWLAGLLCPDPRPLELALLILGVGLLDFCGQVCFTPLEAL 146
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 pfiwalslgillslflipragwlagllcpdprplelallilgvglldfcgqvcftpleal 180

Qy     147 LSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPAIDWDTSALAPYLGTQEECLFGLLTLI 206
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 lsdlfrdpdhcrqaysvyafmislggclgyllpaidwdtsalapylgtqeeclfglltli 240

Qy     207 FLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPHCCPCRARLAFRNLGALLPRLHQLCCR 266
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 fltcvaatllvaeeaalgptepaeglsapslsphccpcrarlafrnlgallprlhqlccr 300

Qy     267 MPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGLYQGVPRAEPGTEARRHYDEGVRMGSL 326
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 mprtlrrlfvaelcswmalmtftlfytdfvgeglyqgvpraepgtearrhydegvrmgsl 360

Qy     327 GLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVAAFPVAAGATCLSHSVAVVTASAALTG 386
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 glflqcaislvfslvmdrlvqrfgtravylasvaafpvaagatclshsvavvtasaaltg 420

Qy     387 FTFSALQILPYTLASLYHREKQVFLPKYRGDTGGASSEDSLMTSFLPGPKPGAPFPNGHV 446
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 ftfsalqilpytlaslyhrekqvflpkyrgdtggassedslmtsflpgpkpgapfpnghv 480

Qy     447 GAGGSGLLPPPPALCGASACDVSVRVVVGEPTEARVVPGRGICLDLAILDSAFLLSQVAP 506
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 gaggsgllppppalcgasacdvsvrvvvgeptearvvpgrgicldlaildsafllsqvap 540

Qy     507 SLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFATQVVFDKSDLAKYSA 553
           ||||||||||||||||||||||||||||||||||||||||||||||
Db     541 slfmgsivqlsqsvtaymvsaaglglvaiyfatqvvfdksdlakysa 587


RESULT  13
AAU69875
ID   AAU69875 standard; Protein; 371 AA.
XX
AC   AAU69875;
XX
DT   30-JAN-2002  (first entry)
XX
DE   Human prostate cDNA encoded protein #69.
XX
KW   Human; prostate cancer; cytostatic; immunostimulant; tumour; immunogen.
XX
OS   Homo sapiens.
XX
PN   WO200173032-A2.
XX
PD   04-OCT-2001.
XX
PF   27-MAR-2001; 2001WO-US09919.
XX
PR   27-MAR-2000; 2000US-0536857.
PR   09-MAY-2000; 2000US-0568100.
PR   12-MAY-2000; 2000US-0570737.
PR   13-JUN-2000; 2000US-0593793.
PR   27-JUN-2000; 2000US-0605783.
PR   10-AUG-2000; 2000US-0636215.
PR   29-AUG-2000; 2000US-0651236.
PR   06-SEP-2000; 2000US-0657279.
PR   02-OCT-2000; 2000US-0679426.
PR   10-OCT-2000; 2000US-0685166.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Xu J,  Dillon DC,  Mitcham JL,  Harlocker SL,  Jiang Y,  Kalos MD;
PI   Fanger GR,  Retter MW,  Stolk JA,  Day CH,  Vedvick TS,  Carter D;
PI   Li SX,  Wang A,  Skeiky YAW,  Hepler WT,  Henderson RA;
XX
DR   WPI; 2001-639232/73.
DR   N-PSDB; AAS64038.
XX
PT   New human prostate-specific polypeptides and polynucleotides useful for
PT   the diagnosis and treatment of cancer, especially prostate cancer -
XX
PS   Claim 2; Page 487-488; 579pp; English.
```

```
XX
CC   The invention relates to isolated prostate-specific
CC   polynucleotides, polypeptides, fusion proteins of the polypeptides,
CC   antibodies raised against the polypeptides (or antigenic epitopes
CC   derived from them) and antigen-presenting cells expressing the
CC   polypeptides. The antibodies are useful for detecting the presence of
CC   cancer, especially prostate cancer. The polypeptides, polynucleotides and
CC   the antigen-presenting cells are useful for stimulating and/or expanding
CC   T cells specific for a tumour protein, and for inhibiting the development
CC   of cancer especially prostate cancer. Compositions comprising the
CC   polynucleotide and/or polypeptide are useful for stimulating an immune
CC   response, and for treating cancer. The oligonucleotide is useful for
CC   detecting cancer. The present sequence is a prostate specific
CC   polypeptide of the invention.
XX
SQ   Sequence   371 AA;


  Query Match             49.5%;  Score 1417.5;  DB 22;  Length 371;
  Best Local Similarity   90.0%;  Pred. No. 3e-129;
  Matches  269;  Conservative    6;  Mismatches   11;  Indels   13;  Gaps    1;

Qy      22 NLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIGPVLGLVCVPLLGSASDHWRGR 81
           :|:   | : |:  :|:   ||             |||||||||||||||||||||||
Db       9 slvplplalylsqplthttsll------------agigpvlglvcvpllgsasdhwrgr 55

Qy      82 YGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDPRPLELALLILGVGLLDFCGQVCFT 141
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      56 ygrrrpfiwalslgillslflipragwlagllcpdprplelallilgvglldfcgqvcft 115

Qy     142 PLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPAIDWDTSALAPYLGTQEECLFG 201
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     116 pleallsdlfrdpdhcrqaysvyafmislggclgyllpaidwdtsalapylgtqeeclfg 175

Qy     202 LLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPHCCPCRARLAFRNLGALLPRLH 261
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     176 lltlifltcvaatllvaeeaalgptepaeglsapslsphccpcrarlafrnlgallprlh 235

Qy     262 QLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGLYQGVPRAEPGTEARRHYDEG 320
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     236 qlccrmprtlrrlfvaelcswmalmtftlfytdfvgeglyqgvpraepgtearrhydeg 294


RESULT  14
AAM01230
ID   AAM01230 standard; Protein; 371 AA.
XX
AC   AAM01230;
XX
DT   04-OCT-2001  (first entry)
XX
DE   P553S splice variant P553S-14 amino acid #2.
XX
KW   Human; prostate cancer; prostate-specific; diagnosis; vaccine;
KW   cytostatic; gene therapy; metastasis.
XX
OS   Homo sapiens.
XX
PN   WO200151633-A2.
XX
PD   19-JUL-2001.
XX
PF   16-JAN-2001; 2001WO-US01574.
XX
PR   14-JAN-2000; 2000US-0483672.
XX
(CORI-) CORIXA CORP.

Dillon DC,  Mitcham JL,  Harlocker SL,  Jiang Y,  Reed SG;
MD,  Fanger GR,  Day CH,  Retter MW,  Stolk JA,  Skeiky YAW;
Meagher MJ;

/45.
```

```
XX
PT   New polynucleotide encoding a prostate-specific protein, for
PT   diagnosing, monitoring and treating prostate cancer in a patient and
PT   for use in vaccines -
XX
PS   Claim 2; Page 464-466; 543pp; English.
XX
CC   The present invention describes polynucleotide sequences (I) which encode
CC   prostate-specific proteins (II). (I) and (II) have cytostatic activity,
CC   and can be used in vaccine production and gene therapy. (I), (II),
CC   antibodies to (II), fusion proteins comprising (II), and isolated
CC   T cells prepared using (I) or (II) are used treat cancer in a patient.
CC   (I) and the antibodies are also used in the detection of cancer in a
CC   patient. The cancer that is diagnosed or treated is particularly
CC   prostate cancer. (I) and (II) can be used in vaccines. The antibodies or
CC   (I) can be used for monitoring the progression of cancer in a patient.
CC   (I) and (II) can also be used to improve diagnostic and therapeutic
CC   methods for prostate cancer. They can indicate the level of metastasis
CC   as well as the prostate volume. AAH93357 to AAH93944 and AAM01115 to
CC   AAM01318 represent polynucleotide and amino acid sequences used in the
CC   exemplification of the present invention.
XX
SQ   Sequence   371 AA;


  Query Match             49.5%;  Score 1417.5;  DB 22;  Length 371;
  Best Local Similarity   90.0%;  Pred. No. 3e-129;
  Matches  269;  Conservative    6;  Mismatches   11;  Indels   13;  Gaps    1;

Qy      22 NLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIGPVLGLVCVPLLGSASDHWRGR 81
           :|:   | : |:  :|:   ||             |||||||||||||||||||||||
Db       9 slvplplalylsqplthttsll------------agigpvlglvcvpllgsasdhwrgr 55

Qy      82 YGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDPRPLELALLILGVGLLDFCGQVCFT 141
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      56 ygrrrpfiwalslgillslflipragwlagllcpdprplelallilgvglldfcgqvcft 115

Qy     142 PLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPAIDWDTSALAPYLGTQEECLFG 201
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     116 pleallsdlfrdpdhcrqaysvyafmislggclgyllpaidwdtsalapylgtqeeclfg 175

Qy     202 LLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPHCCPCRARLAFRNLGALLPRLH 261
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     176 lltlifltcvaatllvaeeaalgptepaeglsapslsphccpcrarlafrnlgallprlh 235

Qy     262 QLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGLYQGVPRAEPGTEARRHYDEG 320
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     236 qlccrmprtlrrlfvaelcswmalmtftlfytdfvgeglyqgvpraepgtearrhydeg 294


RESULT  15
AAE01362
ID   AAE01362 standard; Protein; 371 AA.
XX
AC   AAE01362;
XX
DT   18-JUL-2001  (first entry)
XX
DE   Human gene 11 encoded secreted protein HWBAR14, SEQ ID NO:84.
XX
KW   Human; secreted protein; proliferative disorder; cancer; tumour; asthma;
KW   foetal abnormality; developmental abnormality; haematopoietic disorder;
KW   immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
KW   Parkinson's disease; cognitive disorder; schizophrenia; skin disorder;
KW   psoriasis; sepsis; diabetes; atherosclerosis; cardiovascular disorder;
KW   inflammation; neurological disorder; Alzheimer's disease; food additive;
KW   angiogenic disorder; kidney disorder; gastrointestinal disorder; allergy;
KW   pregnancy-related disorder; endocrine disorder; infection; wound healing;
KW   cell culture; chemotaxis; vulnerary; binding partner identification;
KW   gene therapy.
XX
OS   Homo sapiens.
XX
```

```
FH   Key             Location/Qualifiers
FT   Peptide         1..48
FT                   /label= Signal_peptide
FT   Protein         49..371
FT                   /label= Mature_human_secreted_protein
FT   Misc-difference 20
FT                   /label= Unknown
FT                   /note= "Encoded by TNC"
XX
PN   WO200134629-A1.
XX
PD   17-MAY-2001.
XX
PF   08-NOV-2000; 2000WO-US30654.
XX
PR   12-NOV-1999;   99US-0164835.
PR   27-JUL-2000; 2000US-0221142.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Ruben SM,  Komatsoulis GA,  Wei P,  Fiscella M,  Baker KP;
XX
DR   WPI; 2001-308779/32.
DR   N-PSDB; AAD05230.
XX
PT   New nucleic acid encoding one of 21 human secreted proteins for
PT   diagnosing, preventing, treating or ameliorating medical conditions,
PT   such as autoimmune disease and cancer, and used as a food additive or
PT   preservative -
XX
PS   Claim 11; Page 438-440; 490pp; English.
XX
CC   AAD05220-AAD05282 represent cDNAs corresponding to 21 human secreted
CC   protein genes, and AAE01352-AAE01413 represent the proteins they encode.
CC   AAE01415-AAE01433 represent human secreted protein fragments or variants.
CC   The secreted proteins and their genes are useful for preventing, treating
CC   or ameliorating medical conditions, e.g., by protein or gene therapy.
CC   Pathological conditions can be diagnosed by determining the amount of the
CC   new protein in a sample or by determining the presence of mutations in
CC   the new genes. Specific uses are described for each of the 21 genes,
CC   based on the tissues in which they are most highly expressed, and include
CC   developing products for the diagnosis or treatment of proliferative
CC   disorders, cancer, tumours, foetal and developmental abnormalities,
CC   haematopoietic disorders, diseases of the immune system, AIDS, autoimmune
CC   diseases (e.g., rheumatoid arthritis), inflammation, allergies,
CC   neurological disorders (e.g., Alzheimer's disease, Parkinson's disease),
CC   cognitive disorders, schizophrenia, asthma, skin disorders (e.g.,
CC   psoriasis), sepsis, diabetes, atherosclerosis, cardiovascular disorders,
CC   angiogenic disorders, kidney disorders, gastrointestinal disorders,
CC   pregnancy-related disorders, endocrine disorders, and infections. The
CC   proteins can also be used to aid wound healing and epithelial cell
CC   proliferation, to prevent skin aging due to sunburn, to maintain organs
CC   before transplantation, for supporting cell culture of primary tissues,
CC   to regenerate tissues, to identify their cognate ligands or binding
CC   partners, and in chemotaxis, and can be used as a food additive or
CC   preservative to modify storage properties. Antibodies specific for a
CC   protein of the invention can be used in alleviating symptoms associated
CC   with the disorders mentioned above, and in diagnostic immunoassays e.g.,
CC   radioimmunoassay or enzyme linked immunosorbent assay (ELISA).
CC   The present sequence represents a human secreted protein of
CC   the invention.
XX
SQ   Sequence   371 AA;
```

```
  Query Match             49.5%;  Score 1416;  DB 22;  Length 371;
  Best Local Similarity   98.5%;  Pred. No. 4.1e-129;
  Matches  263;  Conservative    3;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      54 TMVLGIGPVLGLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLL 113
           ::: |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      28 sllagigpvlglvcvpllgsasdhwrgrygrrrpfiwalslgillslflipragwlagll 87

Qy     114 CPDPRPLELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGC 173
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      88 cpdprplelallilgvglldfcgqvcftpleallsdlfrdpdhcrqaysvyafmislggc 147

Qy     174 LGYLLPAIDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLS 233
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     148 lgyllpaidwdtsalapylgtqeeclfglltlifltcvaatllvaeeaalgptepaegls 207

Qy     234 APSLSPHCCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYT 293
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     208 apslsphccpcrarlafrnlgallprlhqlccrmprtlrrlfvaelcswmalmtftlfyt 267

Qy     294 DFVGEGLYQGVPRAEPGTEARRHYDEG 320
           |||||||||||||||||||||||||||
Db     268 dfvgeglyqgvpraepgtearrhydeg 294
```

Search completed: June 26, 2002, 13:43:01
Job time: 4099 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: June 26, 2002, 13:37:17 ; Search time 87.43 Seconds
(without alignments)
1094.203 Million cell updates/sec

Title: US-09-838-785-2
Perfect score: 2861
Sequence: 1 MVQRLWVSRLLRHRKAQLLL.....AIYFATQVVFDKSDLAKYSA 553

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_19:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2861	100.0	553	4	Q96JT2	Q96jt2 homo sapien
2	2798	97.8	553	6	Q95KI5	Q95ki5 macaca fasc
3	2531	88.5	501	6	Q95KC5	Q95kc5 macaca fasc
4	519.5	18.2	599	5	Q9VSV1	Q9vsv1 drosophila
5	456	15.9	576	13	Q90Z74	Q90z74 oryzias lat
6	347.5	12.1	515	10	O65803	O65803 daucus caro
7	347.5	12.1	515	10	Q9FNR6	Q9fnr6 daucus caro
8	346.5	12.1	501	10	Q9SQK6	Q9sqk6 vitis vinif
9	346.5	12.1	515	10	Q9SLN7	Q9sln7 daucus caro
10	346	12.1	612	10	Q9SQK5	Q9sqk5 vitis vinif
11	338	11.8	512	10	Q9S7Z5	Q9s7z5 apium grave
12	337.5	11.8	523	10	O04077	O04077 vicia faba
13	336.5	11.8	501	10	Q9SP63	Q9sp63 vitis vinif
14	333	11.6	512	10	Q9ZTB9	Q9ztb9 apium grave
15	330	11.5	516	10	Q43653	Q43653 solanum tub
16	329.5	11.5	524	10	Q9XHL6	Q9xhl6 pisum sativ
17	328.5	11.5	506	10	Q944W2	Q944w2 oryza sativ
18	328.5	11.5	594	10	O80605	O80605 arabidopsis
19	327.5	11.4	506	10	Q948L0	Q94810 oryza sativ
20	325.5	11.4	604	10	Q9FVL6	Q9fvl6 lycopersico
21	323.5	11.3	512	10	Q39231	Q39231 arabidopsis
22	323.5	11.3	512	10	O80550	O80550 arabidopsis
23	321.5	11.2	502	10	Q9SP14	Q9sp14 alonsoa mer
24	317.5	11.1	507	10	Q40583	Q40583 nicotiana t
25	317.5	11.1	530	10	Q9M535	Q9m535 euphorbia e
26	316.5	11.1	500	10	Q9FVJ6	Q9fvj6 lycopersico
27	313	10.9	521	10	Q9SXM0	Q9sxm0 zea mays (m
28	311	10.9	533	10	Q41152	Q41152 ricinus com
29	307.5	10.7	488	10	Q9FV92	Q9fv92 solanum tub
30	306	10.7	523	10	Q9M422	Q9m422 hordeum vul
31	305.5	10.7	501	10	O65929	O65929 daucus caro
32	302.5	10.6	510	10	Q9SP15	Q9sp15 asarina bar
33	301.5	10.5	429	10	Q40167	Q40167 lycopersico
34	299.5	10.5	505	10	Q9SQK4	Q9sqk4 vitis vinif
35	296.5	10.4	510	10	Q9M3R4	Q9m3r4 arabidopsis
36	294.5	10.3	510	10	Q9FE59	Q9fe59 arabidopsis
37	294	10.3	510	10	Q40938	Q40938 plantago ma
38	294	10.3	523	10	Q39438	Q39438 beta vulgar
39	292.5	10.2	512	10	Q9C8X2	Q9c8x2 arabidopsis
40	292	10.2	537	10	O49838	O49838 oryza sativ
41	292	10.2	538	10	Q9LKH3	Q9lkh3 oryza sativ
42	291.5	10.2	509	10	Q9FNY9	Q9fny9 arabidopsis
43	291	10.2	618	4	Q9ULU3	Q9ulu3 homo sapien
44	287	10.0	539	10	Q9SM24	Q9sm24 beta vulgar
45	286	10.0	474	10	O04516	O04516 arabidopsis

ALIGNMENTS

RESULT 1
Q96JT2
ID Q96JT2 PRELIMINARY; PRT; 553 AA.
AC Q96JT2;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE PROSTEIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PROSTATE;
RX MEDLINE=21139094; PubMed=11245466;
RA Xu J., Kalos M., Stolk J.A., Zasloff E.J., Zhang X., Houghton R.L.,
RA Filho A.M., Nolasco M., Badaro R., Reed S.G.;
RT "Identification and characterization of prostein, a novel prostate-
RT specific protein.";
RL Cancer Res. 61:1563-1568(2001).
DR EMBL; AY033593; AAK54386.1; -.
SQ SEQUENCE 553 AA; 59322 MW; 0AFA23FBC742A667 CRC64;

Query Match 100.0%; Score 2861; DB 4; Length 553;
Best Local Similarity 100.0%; Pred. No. 5.4e-213;
Matches 553; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 MVQRLWVSRLLRHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIG 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MVQRLWVSRLLRHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIG 60

Qy      61 PVLGLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDPRPL 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 PVLGLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDPRPL 120

Qy     121 ELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPA 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    121 ELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPA 180

Qy    181 IDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPH 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 IDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPH 240

Qy    241 CCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGL 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 CCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGL 300

Qy    301 YQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVA 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 YQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVA 360

Qy    361 AFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 AFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420

Qy    421 ASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 ASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480

Qy    481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFATQ 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFATQ 540

Qy    541 VVFDKSDLAKYSA 553
          |||||||||||||
Db    541 VVFDKSDLAKYSA 553
```

```
RESULT   2
Q95KI5
ID   Q95KI5      PRELIMINARY;      PRT;   553 AA.
AC   Q95KI5;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   HYPOTHETICAL 59.4 KDA PROTEIN.
OS   Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
OC   Cercopithecinae; Macaca.
OX   NCBI_TaxID=9541;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=TEMPORAL LOBE RIGHT;
RA   Osada N., Hida M., Kusuda J., Tanuma R., Iseki K., Hirai M., Terao K.,
RA   Suzuki Y., Sugano S., Hashimoto K.;
RT   "Isolation of full-length cDNA clones from macaque brain cDNA
RT   libraries.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB060851; BAB46871.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   553 AA;  59392 MW;  0718F3A91FB3BF1E CRC64;
```

```
  Query Match             97.8%;  Score 2798;  DB 6;  Length 553;
  Best Local Similarity   98.2%;  Pred. No. 4e-208;
  Matches  542;  Conservative    2;  Mismatches    8;  Indels    0;  Gaps    0;

Qy      1 MVQRLWVSRLLRHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIG 60
          |||||||||||||||||||:|||||||||||||||||||||||||||||||||||||||||
Db      1 MVQRLWVSRLLRHRKAQLLLINLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIG 60

Qy     61 PVLGLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDPRPL 120
          |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 PVLGLVSVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDPRPL 120

Qy    121 ELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPA 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 ELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPA 180
```

```
Qy    181 IDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPH 240
          ||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||  |
Db    181 IDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPAEPAEGLSAPSLPSH 240

Qy    241 CCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGL 300
          |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 CCPCWARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGL 300

Qy    301 YQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVA 360
          |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 YQGVPRAELGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVA 360

Qy    361 AFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420
          ||||||||||||||||||||||||||||||||||||||||||||||:|||||||||| ||
Db    361 AFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHRERQVFLPKYRGDAGG 420

Qy    421 ASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 TSSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480

Qy    481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFATQ 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFATQ 540

Qy    541 VVFDKSDLAKYS 552
          ||||||||||||
Db    541 VVFDKSDLAKYS 552
```

```
RESULT   3
Q95KC5
ID   Q95KC5      PRELIMINARY;      PRT;   501 AA.
AC   Q95KC5;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   HYPOTHETICAL 53.4 KDA PROTEIN.
OS   Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
OC   Cercopithecinae; Macaca.
OX   NCBI_TaxID=9541;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=MEDULLA OBLOGATA;
RA   Osada N., Hida M., Kusuda J., Tanuma R., Iseki K., Hirai M., Terao K.,
RA   Suzuki Y., Sugano S., Hashimoto K.;
RT   "Isolation of full-length cDNA clones from macaque brain cDNA
RT   libraries.";
RL   Submitted (JUN-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB062977; BAB60745.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   501 AA;  53447 MW;  8C554BBD04EE0470 CRC64;
```

```
  Query Match             88.5%;  Score 2531;  DB 6;  Length 501;
  Best Local Similarity   97.8%;  Pred. No. 1.5e-187;
  Matches  489;  Conservative    2;  Mismatches    9;  Indels    0;  Gaps    0;

Qy     53 MTMVLGIGPVLGLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGL 112
          |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MTMVLGIGPVLGLVSVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGL 60

Qy    113 LCPDPRPLELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGG 172
          |||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||
Db     61 LCPDPRPLELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYTFMISLGG 120

Qy    173 CLGYLLPAIDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGL 232
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db    121 CLGYLLPAIDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPAEPAEGL 180

Qy    233 SAPSLSPHCCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFY 292
          |||||  ||||| ||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    181 SAPSLPSHCCPCWARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFY 240

Qy    293 TDFVGEGLYQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTR 352
          |||||||||||||||| ||||||||||||||||||||||||||||||:||||||||||
Db    241 TDFVGEGLYQGVPRAELGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVVDRLVQRFGTR 300

Qy    353 AVYLASVAAFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLP 412
          |||||||||||||||||||||||||||||||||||||||||||||||||||||:|||||
Db    301 AVYLASVAAFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHRERQVFLP 360

Qy    413 KYRGDTGGASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRV 472
          ||||| || ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 KYRGDAGGTSSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRV 420

Qy    473 VVGEPTEARVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGL 532
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 VVGEPTEARVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGL 480

Qy    533 VAIYFATQVVFDKSDLAKYS 552
          ||||||||||||||||||||
Db    481 VAIYFATQVVFDKSDLAKYS 500
```

RESULT 4
Q9VSV1
ID Q9VSV1 PRELIMINARY; PRT; 599 AA.
AC Q9VSV1;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE CG4484 PROTEIN.
GN CG4484.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL; AE003552; AAF50310.1; -.
DR FlyBase; FBgn0035968; CG4484.
SQ SEQUENCE 599 AA; 66057 MW; C5381D334CFBF2EB CRC64;

Query Match 18.2%; Score 519.5; DB 5; Length 599;
Best Local Similarity 23.7%; Pred. No. 4.7e-32;
Matches 150; Conservative 100; Mismatches 201; Indels 183; Gaps 14;

```
Qy     12 RHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIGPVLGLVCVPLL 71
          |  : ::  :: :   :|   ||  ::| |:||::||: | |:|  |: |::|    |||
Db     46 RKTRFEMFRLSAIAMAIEFAYAAETSFVSPILLQIGVDHKHMSMTWGLSPLIGFFMSPLL 105

Qy     72 GSASDHWRGRYGRRRPFIWALSLGILLSLFLIP-------------------------R 105
          || ||  : |:|||| |  || ||:  | |:|
Db    106 GSISDRCKLRWGRRRPIISILSFGIMCGLILVPYGKDLGLLLGDAGYTYAESALNFTSSS 165

Qy    106 AGWLAGLLC------PDPRPLELA--LLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHC 157
          | :| |:          |     : |  | |||: ||||    | ||    | |:   |:
Db    166 GGSVAALVSGEATTGPSASDYKFAVILTILGMVLLDFDADTCQTPARTYLLDMC-VPEEQ 224

Qy    158 RQAYSVYAFMISLGGCLGYLLPAIDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATL-- 215
          :| :::|     || :|| :  :||:|: :  ::|     :| |:|:||  |   |:
Db    225 PKAMTMFALFAGFGGTIGYAIGGVDWETTHIGSFMGGNIPTVFTLVTIIFAVCYLITVTT 284

Qy    216 -------LVAEEAALGP-----------------------------------TEPAEG 231
                |: ::  | |                                    |  :|
Db    285 FREIPLPLIEQDELLRPLSEQAIKKELKKKNNTIYYIQETTQLELQMASDDPKRLEALQG 344

Qy    232 LSAPSLSP----------------HCCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLF 275
                ||                :  |   :   :::             || ::| |
Db    345 SYQNGYSPAVEKQGKSQDLETQSDYDAPVSLKAYLKSIFI-----------MPYSMRMLA 393

Qy    276 VAELCSWMALMTFTLFYTDFVGEGLYQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAIS 335
          :  |  ||  :|: |::|||||| :: | | | | :||  :|: ||| |  |: :
Db    394 LTNLFCWMGHVTYCLYFTDFVGEAVFHGDPTAAPNSEAALNYEAGVRFGCWGMAIYAFSC 453

Qy    336 LVFSLVMDRLVQRFGTRAVYLASVAAFPV---AAGATCLSHSVAVVTASAALTGFTFSAL 392
          ::|| : :|:: |||:|||:: :  : :     |     | | : ||   |  : :
Db    454 SIYSLSVTKLMKWFGTKAVYISGMIYYGIGMLVLGLWPTKWGVLVFSTSA---GILYGTI 510

Qy    393 QILPYTLASLYHREKQVFLPKYRGDTGGASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSG 452
          :|: | : ||  |  |  |                                  ||
Db    511 FTVPFILVARYH-AKNCFSIK----------------------------NG-------- 532

Qy    453 LLPPPPALCGASACDVSVRVVVGEPTEARVVP---GRGICLDLAILDSAFLLSQVAPSLF 509
                                          :||    ||:  |:||: |   ::|:  ||
Db    533 ---------------------------EIVPLKQARGLGTDVAIISSMVFIAQLIVSLS 564

Qy    510 MGSIVQLSQSVTAYMVSAAGLGLVAIYFATQVVF 543
          :| :|    :  | : ::  |  :|   |  |::
Db    565 VGPLVSWMDTTCAVLYASTFLSFLAAIAAMFVLY 598
```

RESULT 5
Q90Z74
ID Q90Z74 PRELIMINARY; PRT; 576 AA.
AC Q90Z74;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE MEMBRANE-ASSOCIATED TRANSPORTER PROTEIN B.
OS Oryzias latipes (Medaka fish).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
OC Beloniformes; Adrianichthyidae; Oryziinae; Oryzias.

```
OX   NCBI_TaxID=8090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=HNI;
RX   MEDLINE=21372467; PubMed=11479596;
RA   Fukamachi S., Shimada A., Shima A.;
RT   "Mutations in the gene encoding B, a novel transporter protein, reduce
RT   melanin content in medaka.";
RL   Nat. Genet. 28:381-385(2001).
DR   EMBL; AF332510; AAK77024.1; -.
SQ   SEQUENCE   576 AA;  63411 MW;  13A5C9828ABEE894 CRC64;
```

```
  Query Match             15.9%;  Score 456;  DB 13;  Length 576;
  Best Local Similarity   24.0%;  Pred. No. 3.6e-27;
  Matches  144;  Conservative   95;  Mismatches  198;  Indels  164;  Gaps   17;

Qy      12 RHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIGPVLGLVCVPLL 71
           |  : :|:| ::: || | | |     :| |:|| ||:     ::|  | |:|| :  |::
Db      56 RRSRGRLILHSMVMFGREFCYAVEAAFVTPVLLSVGLPRSLYSLVWLISPILGFLLQPII 115

Qy      72 GSASDHWRGRYGRRRPFIWALSLGIL----LSLFL-------------IPRAGWLAGLLC 114
           |||||: |  :|||||:|  | ||||    ||:||                |: |
Db     116 GSASDYCRSSWGRRRPYI--LVLGILMLVGLSMFLNGDAVVSELVSDRSSRSTW------ 167

Qy     115 PDPRPLELALLILGVGLLDFCGQVCFTPLEALLSDL--FRDPDHCRQAYSVYAFMISLGG 172
                : ::: || | ||          |::| | |:  ::| :   :    :|    |||
Db     168 ------AIVVVMFGVVLFDFAADFIDGPIKAYLFDVCSYQDKE---RGLHYHALFTGLGG 218

Qy     173 CLGYLLPAIDWDTSALAPYLGTQEECLF-------GLLTLIFLTCV-------------- 211
           |||: |:||  | |   ||:: : ::       |:  :: |  :
Db     219 ACGYLVGAMDWGHSVLGRLLGSEYQVIYFFSALTWGVFLIVHLFSIPEKPLAKVPSESSA 278

Qy     212 -AATLLVAEEA----ALG--PTEPAEGLSAPSLSPHCCPC-------------------- 244
           :|  |:   :    |||  |  |    |:| : |
Db     279 SSALRLLGPHSNGYGALGKEPVSPVIPTSSPEIRPRSYSALGERPRSFSALGEANSVTSS 338

Qy     245 --------RARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFV 296
                  : :: ||:       | :    ||   | | :: |  | | :   ||:|||:
Db     339 AKQPIKEDQKKMTFRS-------LMKAIFNMPNHYRFLCISHLLGWAAFLCNMLFFTDFM 391

Qy     297 GEGLYQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYL 356
           |: :|:| | ||  : |   |: || :|  || :    | ::| |   |:   | : :|
Db     392 GQIVYRGNPYAEHNSTAYITYERGVEVGCWGLCINAVSSALYSYVQRFLLPYIGLKGLYF 451

Qy     357 ASVAAFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRG 416
             | :    |  |       :: |   | |  :|: | : | ||::  : |  |
Db     452 MGYFVFGMGTSLIGLFPEVIATLILCSVFGVMSSTLYTIPFNLIAEYQREEEEQV-KLEG 510

Qy     417 DTGGASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGE 476
                                    |:   |:|:       | |  | | :
Db     511 --------------------------GNESPRGTGM------DCAALTCMVQL------ 531

Qy     477 PTEARVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIY 536
             |::: | |                        :|::|  : ||   ::||: : |:
Db     532 ---AQIIVGAG--------------------LGALVNAAGSVIVVVLSASSISLIGCI 566

Qy     537 F 537
           |
Db     567 F 567
```

```
RESULT   6
O65803
ID   O65803      PRELIMINARY;      PRT;   515 AA.
AC   O65803;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
     SUCROSE/H+ SYMPORTER.
     S
     O          arota (Carrot).
     O          iridiplantae; Streptophyta; Embryophyta; Tracheophyta;
```

```
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC   Asteridae; euasterids II; Apiales; Apiaceae; Daucus.
OX   NCBI_TaxID=4039;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. NAMTAISE; TISSUE=ROOT;
RX   MEDLINE=99063785; PubMed=9847123;
RA   Shakya R., Sturm A.;
RT   "Characterization of source- and sink-specific sucrose/H+ symporters
RT   from carrot.";
RL   Plant Physiol. 118:1473-1480(1998).
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.
DR   EMBL; Y16768; CAA76369.1; -.
DR   InterPro; IPR003662; sub_transporter.
DR   Pfam; PF00083; sugar_tr; 1.
KW   Transmembrane.
SQ   SEQUENCE   515 AA;  54424 MW;  897B90657C9E243C CRC64;
```

```
  Query Match             12.1%;  Score 347.5;  DB 10;  Length 515;
  Best Local Similarity   25.9%;  Pred. No. 7.7e-19;
  Matches  131;  Conservative   80;  Mismatches  215;  Indels   79;  Gaps   17;

Qy      17 QLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIGPVLGLVCVPLLGSASD 76
           :|:||  :  |::   |  :: : | :  :|:  |:   :   ||: |::  |::|  ||
Db      34 KLVLVAAIAAGVQFGWALQLSLLTPYVQLLGIPHKWAAYIWLCGPISGMLVQPIVGYYSD 93

Qy      77 HWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGL-------LCPDPRPLELALLILGV 129
           | :  :|||||||| : :  : :|: ||   |: | :       :    :|  : : ::|
Db      94 HCQSSFGRRRPFIASGAGCVAISVILI---GFAADISYKAGDDMSKTLKPRAVTVFVIGF 150

Qy     130 GLLDFCGQVCFTPLEALLSDLFR-DPDHCRQAYSVYAFMISLGGCLGY----------LL 178
           :||    :   |  |||:||   |    | | : |:| :::|  |||          |
Db     151 WILDVANNMLQGPCRALLADLCSGDTRRMRSANAFYSFFMAVGNILGYAAGSYNNLYKLF 210

Qy     179 PAIDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLS 238
           |    | |   |    : |    : |: :  | |  :| | :  || : |:    |
Db     211 PFS--KTHACDLYCANLKSCFIISIALLIIITVVALSVVRENS--GPPDDADAAEEP--- 263

Qy     239 PHCCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGE 298
               |   ::     | ||  |  |    || :  | :    :|:|   | || ||:|
Db     264 ----PSSGKIPV--FGELLGALKDL----PRPMLLLLIVTCLNWIAWFPFILFDTDWMGR 313

Qy     299 GLYQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLAS 358
           :| |   |  |    : ||:||| |:||| |   :  : |: :: ||:  |   : |
Db     314 EIYGGT--AGQG----KLYDQGVRAGALGLLLNSVVLGLTSIAVEYLVRGVGGVKI-LWG 366

Qy     359 VAAFPVAAGATCL-----------SHSV--------AVVTASA----ALTGFTFSALQIL 395
           | :| |                  ||         | | | |    :: |   |    :
Db     367 FVNFILAIGLVMTVVVSKVAQHQREHSANGQLLPPSAGVKAGALSLFSILGIPLSITYSI 426

Qy     396 PYTLASLYHREKQVFLPKYRGDTGGASSEDSLMTSFLPGPKPG-------APFPNGHVGA 448
           |: |||:|          |    |     :: | | ||             |  | : |
Db     427 PFALASIYSSGSGAGQGLSLGVLNLAIVVPQMIVSVLAGPFDSLFGGGNLPAFVVGAISA 486

Qy     449 GGSGLLP----PPPALCGASACDVS 469
           ||:|     | |:   ||   :|
Db     487 AISGVLAIVLLPKPSKDAASKLSLS 511
```

```
RESULT   7
Q9FNR6
ID   Q9FNR6      PRELIMINARY;      PRT;   515 AA.
AC   Q9FNR6;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   SUCROSE/PROTON SYMPORTER.
GN   SUT2.
OS   Daucus carota (Carrot).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
```

```
OC   Asteridae; euasterids II; Apiales; Apiaceae; Daucus.
OX   NCBI_TaxID=4039;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. NANTAISE;
RA   Sturm A.;
RL   Submitted (DEC-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. NANTAISE;
RA   Shakya R.;
RL   Thesis (2000), Department of Department of Botany,
RL   University of Basel, Basel, Switzerland.
DR   EMBL; AJ303199; CAC19689.1; -.
SQ   SEQUENCE   515 AA;  54425 MW;  9580978C726523DC CRC64;

  Query Match             12.1%;  Score 347.5;  DB 10;  Length 515;
  Best Local Similarity   25.9%;  Pred. No. 7.7e-19;
  Matches  131;  Conservative   80;  Mismatches  215;  Indels   79;  Gaps   17;

Qy      17 QLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIGPVLGLVCVPLLGSASD 76
           :|:||  :  |::   |  :: : | :  :|:  |:   :   ||: |::  |::|  ||
Db      34 KLVLVAAIAAGVQFGWALQLSLLTPYVQLLGIPHKWAAYIWLCGPISGMLVQPIVGYYSD 93

Qy      77 HWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGL-------LCPDPRPLELALLILGV 129
           | :  :|||||||| : :  : :|: ||   |: | :       :   :|  : : ::|
Db      94 HCQSSFGRRRPFIASGAGCVAISVILI---GFAADISYKAGDDMSKTLKPRAVTVFVIGF 150

Qy     130 GLLDFCGQVCFTPLEALLSDLFR-DPDHCRQAYSVYAFMISLGGCLGY----------LL 178
           :||    :   |  |||:||   |    | | : |:| :::|  |||          |
Db     151 WILDVANNMLQGPCRALLADLCSGDTRRMRSANAFYSFFMAVGNILGYAAGSYNNLYKLF 210

Qy     179 PAIDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLS 238
           |     | |   |    : |    : |: :  | |  :| | :  || : |:    |
Db     211 PFS--KTHACDLYCANLKSCFIISIALLIIITVVALSVVRENS--GPPDDADAAEEP--- 263

Qy     239 PHCCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGE 298
           |    ::     | ||  | |    || :  | :    :|:|   | || ||::|
Db     264 ----PSSGKIPV--FGELLGALKDL----PRPMLLLLIVTCLNWIAWFPFILFDTDWMGR 313

Qy     299 GLYQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLAS 358
           :| |   | |    : ||:||| |:||| |   :  : |: :: ||:  |   : |
Db     314 EIYGGT--AGKG----KLYDQGVRAGALGLLLNSVVLGLTSIAVEYLVRGVGGVKI-LWG 366

Qy     359 VAAFPVAAGATCL-----------SHSV--------AVVTASA----ALTGFTFSALQIL 395
             | :| |                ||          | | | |    ::  |    |    :
Db     367 FVNFILAIGLVMTVVVSKVAQHQREHSANGQLLPPSAGVKAGALSLFSILGIPLSITYSI 426

Qy     396 PYTLASLYHREKQVFLPKYRGDTGGASSEDSLMTSFLPGPKPG-------APFPNGHVGA 448
           |: |||:|              |    |     :: | | ||              |  | : |
Db     427 PFALASIYSSGSGAGQGLSLGVLNLAIVVPQMIVSVLAGPFDSLFGGGNLPAFVVGAISA 486

Qy     449 GGSGLLP----PPPALCGASACDVS 469
             ||:|     | |:    ||   :|
Db     487 AISGVLAIVLLPKPSKDAASKLSLS 511


RESULT   8
Q9SQK6
ID   Q9SQK6      PRELIMINARY;      PRT;   501 AA.
AC   Q9SQK6;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   PUTATIVE SUCROSE TRANSPORTER.
GN   VVSUC11.
OS   Vitis vinifera (Grape).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Vitaceae;
OC   Vitis.
OX   NCBI_TaxID=29760;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SHIRAZ; TISSUE=FRUIT;
RA   Davies C., Wolf T., Robinson S.P.;
RT   "Three putative sucrose transporters are differentially expressed in
RT   grapevine tissues.";
RL   Plant Sci. 147:93-100(1999).
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.
DR   EMBL; AF021808; AAF08329.1; -.
DR   InterPro; IPR003662; sub_transporter.
DR   Pfam; PF00083; sugar_tr; 1.
KW   Transmembrane.
SQ   SEQUENCE   501 AA;  53938 MW;  4D0D4DE2EF2F4BA8 CRC64;

  Query Match             12.1%;  Score 346.5;  DB 10;  Length 501;
  Best Local Similarity   27.0%;  Pred. No. 8.9e-19;
  Matches  153;  Conservative   83;  Mismatches  197;  Indels  133;  Gaps   22;

Qy      17 QLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIGPVLGLVCVPLLGSASD 76
           :|| |  :  |::   |  :: : | : |:|:   : :::   ||: ||:  ||:|  ||
Db      29 RLLRVASVACGIQFGWALQLSLLTPYVQELGIPHAWSSIIWLCGPLSGLLVQPLVGHLSD 88

Qy      77 HWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLL--CPDPRPLELALLILGVGLLDF 134
             |:|||||| | :  |:::: :|  :  : |||    | ||  :|  ::|  |||
Db      89 RCNSRFGRRRPFIVAGATSIVVAVLIIGFSADIGGLLGDGADRRPRAVATFVVGFWLLDV 148

Qy     135 CGQVCFTPLEALLSDLFRDPDH--CRQAYSVYAFMISLGGCLGYLLPA------IDWDTS 186
           |   |  |||:||  : ||   | | : ::  |::|  ||:   :       | | ||
Db     149 ANNVTQGPCRALLADL-TEKDHRRTRVANAYFSLFIAVGNVLGFATGSYSGWFRIFWFTS 207

Qy     187 ALAPYLGTQEECL----FGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPHCC 242
           :      :|       || :||   :| |  ::  ||         ||: | | |
Db     208 TSS----CNADCANLKSAFLLDIIF---IAITTYISITAA-----QELPLSSSSRSTHIS 255

Qy     243 -----PCRARLAFRNLGALLPRLHQLCCRMPRTLRR--------LFVAELCSWMALMTFT 289
                 |: ||         | :|     |||          |||  | :|: |: |
Db     256 EEMAESTHAQEAF---------LWELF----GTLRYLSGSIWIILFVTAL-TWIGLLPFL 301

Qy     290 LFYTDFVGEGLYQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRF 349
           || ||::|  :| | |          ::|: |||||:||| |   :  : |::|::| :::
Db     302 LFDTDWMGREIYGGKP------NEGQNYNTGVRMGALGLMLNSVVLGITSVLMEKLCRKW 355

Qy     350 GTRAVYLASVAAFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQV 409
           |  |:                    | |    |  : ||    | : | :
Db     356 GAGFVW----------------------GLSNILMSLCFLLMLILS---AVVKHMD--- 386

Qy     410 FLPKYRGDTGGASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVS 469
                                ||    |    |:| | |          ||   |   :
Db     387 --------------------FLGHDLP----PSGVVIA----------ALIVFSILGIP 411

Qy     470 VRVVVGEP---TEARVVP-GRGICLDLAILDSAFLLSQVAPSLFMGSIVQL-----SQSV 520
           : :    |      |:  | |  | : :|: | :: ||  ||  |   ||     | |:
Db     412 LAITYSVPYALISTRIESLGLGQGLSMGVLNLAIVIPQVIVSLGSGPWDQLFGGGNSPSL 471

Qy     521 TAYMVSAAGLGLVAIYFATQVVFDKS 546
               |:|   |||||    :   |||
Db     472 AVAAVAAFASGLVAILAIPRSSADKS 497


RESULT   9
Q9SLN7
ID   Q9SLN7      PRELIMINARY;      PRT;   515 AA.
AC   Q9SLN7;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   SUCROSE TRANSPORTER PROTEIN.
GN   CSUT.
OS   Daucus carota (Carrot).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC   Asteridae; euasterids II; Apiales; Apiaceae; Daucus.
```

```
OX   NCBI_TaxID=4039;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=ROOT;
RA   Wu N., Diao F., Zhang L., Huang M.;
RT   "Sucrose transporter protein.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.
DR   EMBL; AB036758; BAA89458.1; -.
DR   InterPro; IPR003662; sub_transporter.
DR   Pfam; PF00083; sugar_tr; 1.
KW   Transmembrane.
SQ   SEQUENCE   515 AA;  54523 MW;  2B548D1D9DFF51AC CRC64;
```

```
  Query Match             12.1%;  Score 346.5;  DB 10;  Length 515;
  Best Local Similarity   27.2%;  Pred. No. 9.1e-19;
  Matches  137;  Conservative   76;  Mismatches  215;  Indels   75;  Gaps   19;

Qy      17 QLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIGPVLGLVCVPLLGSASD 76
           :|:||  :  |::   |  :: : | :  :|:  |:   :   ||: |::  |::|  ||
Db      34 KLVLVAAIAAGVQFGWALQLSLLTPYVQLLGIPHKWAAYIWLCGPISGMLVQPIVGYYSD 93

Qy      77 HWRGRYGRRRPFIWALSLGILLSLFLIPRA---GWLAG-LLCPDPRPLELALLILGVGLL 132
           | :  :|||||| : :  : :|: ||  |   |: ||  :   :|  : : ::|  :|
Db      94 HCQSSFGRRRPFIASGAGCVAISVILIGFAADIGYKAGDDMSKTLKPRAVTVFVIGFWIL 153

Qy     133 DFCGQVCFTPLEALLSDLFR-DPDHCRQAYSVYAFMISLGGCLGY----------LLPAI 181
           |    :  | |||:||   |   | | : |:| :::|  |||          | |
Db     154 DVANNMLQGPCRALLADLCSGDTRRMRSANAFYSFFMAVGNILGYAAGSYNNLYKLFPFS 213

Qy     182 DWDTSALAPYLGTQEEC-LFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPH 240
           | |   |   : | :  :  || :| ||  | |  |    | : |:    |
Db     214 --KTHACDLYCANLKSCFIISIALLIIITVVA--LSVVREKQWSP-DDADAADEP----- 263

Qy     241 CCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGL 300
           |   ::     | ||  |  |    || :  | :    :|:|   | || ||::|  :
Db     264 --PSSGKIPV--FGELLGALKDL----PRPMLLLLIVTCLNWIAWFPFILFDTDWMGREI 315

Qy     301 YQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVA 360
           | |   |  |    : ||:||| ||||| |   :  : |: :: ||:  |   : | |
Db     316 YGGT--AGKG----KLYDQGVRAGSLGLLLNSVVLGLTSIAVEYLVRGVGGVKI-LWGVV 368

Qy     361 AFPVAAGATCL----------SHSV--------AVVTASA----ALTGFTFSALQILPY 397
           | :| |                ||          | | | |    ::  |   |    :|:
Db     369 NFILAIGLVMTVVVSKVAQHQREHSANGQLLPPSAGVKAGALSLFSILGIPLSITYSIPF 428

Qy     398 TLASLYHREKQVFLPKYRGDTGGASSEDSLMTSFLPGPKPG-------APFPNGHVGAGG 450
           |||:|           |    |      :: | | ||              |  | : |
Db     429 ALASIYSSGSGAGQGLSLGVLNLAIVVPQMIVSVLAGPFDSLFGGGNLPAFVVGAISAAI 488

Qy     451 SGLLP----PPPALCGASACDVS 469
           ||:|     | |:   ||   :|
Db     489 SGVLAIVLLPKPSKDAASKLSLS 511
```

```
RESULT  10
Q9SQK5
ID   Q9SQK5      PRELIMINARY;      PRT;   612 AA.
AC   Q9SQK5;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   PUTATIVE SUCROSE TRANSPORTER.
GN   VVSUC12.
OS   Vitis vinifera (Grape).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Vitaceae;
OC   Vitis.
OX   NCBI_TaxID=29760;
RN   [1]
RP   SEQUENCE FROM N.A.
```

```
RC   STRAIN=SHIRAZ; TISSUE=FRUIT;
RA   Davies C., Wolf T., Robinson S.P.;
RT   "Three putative sucrose transporters are differentially expressed in
RT   grapevine tissues.";
RL   Plant Sci. 147:93-100(1999).
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.
DR   EMBL; AF021809; AAF08330.1; -.
DR   InterPro; IPR003662; sub_transporter.
DR   Pfam; PF00083; sugar_tr; 1.
KW   Transmembrane.
SQ   SEQUENCE   612 AA;  65711 MW;  8FC3FBD6AB439078 CRC64;
```

```
  Query Match             12.1%;  Score 346;  DB 10;  Length 612;
  Best Local Similarity   25.4%;  Pred. No. 1.2e-18;
  Matches  125;  Conservative   75;  Mismatches  178;  Indels  114;  Gaps   17;

Qy      18 LLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIGPVLGLVCVPLLGSASDH 77
           |:|  ::  |::   |  :: : | :  :|:|  | : :   ||: |||  | :|  ||
Db      66 LILSCMIAAGVQFGWALQLSLLTPYIQTLGIEHAFSSFIWLCGPITGLVVQPCVGIWSDK 125

Qy      78 WRGRYGRRRPFIWALSLGILLSLFLI---PRAGWLAGLLCPDPRPLE------LALLILG 128
              :||||||| | || | ::: :|      |:| |    | |  :          : :||
Db     126 CSSKYGRRRPFILAGSLMISVAVTIIGFSADIGYLLGDTNMDCRKFKGTRTWAAIIFVLG 185

Qy     129 VGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPAID----WD 184
           :||        |  |||:||   ||    | :::   :::|  ||:   |      |
Db     186 FWMLDLANNTVQGPARALLADL-SGPDQRNSANAIFCSWMAVGNILGFSAGASGHWHRW- 243

Qy     185 TSALAPYLGTQEEC-------LFGLLTLIFLT-CVAATLLVAEEAALGPTEPAE-GLSAP 235
              |:|  :  |          |: ::||| |   ||  |||  |   :|     |||
Db     244 ----FPFLLNKACCEACGNLKAAFLIAVVFLTLCTLVTLYFAEEVPLMAYQPHHLSDSAP 299

Qy     236 SL-SPHCCPCRARLAFRNL---------------------------------------- 253
           | :|       :: | |
Db     300 LLDNPQ------QIGFDNSKSKLDMSAVDNATGNNPESSYEINKNAKHLTPIVQEQNESF 353

Qy     254 ----GALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGLYQGVPRAEP 309
              ||:|  |      :|  :  : :    ||::   | || ||::|  :| | |: :
Db     354 SDGPGAVLVNLLTSLRHLPPAMHSVLLVMALSWLSWFPFFLFDTDWMGREVYHGDPKGD- 412

Qy     310 GTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVY-LASVAAFPVAAGA 368
           | : || ||| |: || |       : | ::: : || | | |: :::   |   ||
Db     413 -ESAVKAYDAGVREGAFGLLLNSVDLGISSFLIEPMCQRMGARLVWAMSNFIVFACMAGT 471

Qy     369 TCLS------------HSV----AVVTASA---ALTGFTFSALQILPYTLASLYHREKQ 408
             :|            |::    |:  ||    || ||  |    :|::: :
Db     472 AIISLVSVNEYITEGIQHAIGENRAIKIASLVVFALLGFPLSITYSVPFSITA------- 524

Qy     409 VFLPKYRGDTGG 420
           :  ||||
Db     525 ----ELTADTGG 532
```

```
RESULT  11
Q9S7Z5
ID   Q9S7Z5      PRELIMINARY;      PRT;   512 AA.
AC   Q9S7Z5;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE   SUCROSE TRANSPORTER SUT2B.
GN   SUT2B OR SUT2A.
OS   Apium graveolens (Celery).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC   Asteridae; euasterids II; Apiales; Apiaceae; Apium.
OX   NCBI_TaxID=4045;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=MATURE LEAF;
RA   Noiraud N., Lemoine R.;
```

```
RT   "Sucrose transporters in celery.";
RL   Submitted (JUL-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.
DR   EMBL; AF167416; AAD45391.1; -.
DR   EMBL; AF167415; AAD45390.1; -.
DR   InterPro; IPR003662; sub_transporter.
DR   Pfam; PF00083; sugar_tr; 1.
KW   Transmembrane.
SQ   SEQUENCE   512 AA;  54426 MW;  2637535216FF1ED2 CRC64;


  Query Match             11.8%;  Score 338;  DB 10;  Length 512;
  Best Local Similarity   25.3%;  Pred. No. 4.1e-18;
  Matches  140;  Conservative   87;  Mismatches  198;  Indels  128;  Gaps   21;

Qy      17 QLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIGPVLGLVCVPLLGSASD 76
           :|:||  :  |::   |  :: : | :  :|:  |:   :   ||: |::  |::|  ||
Db      33 KLILVAAIAAGVQFGWALQLSLLTPYVQLLGIPHKWAAYIWLCGPISGMLVQPIVGYYSD 92

Qy      77 HWRGRYGRRRPFIWALSLGILLSLFLIPRA---GWLAG-LLCPDPRPLELALLILGVGLL 132
             :  :|||||||| : :  : :|: ||  |   |: || :    :|  :   ::|  :|
Db      93 RCQSSFGRRRPFIASGAGCVAISVILIGFAADIGYKAGDDMTKTLKPRAVTGFVIGFWIL 152

Qy     133 DFCGQVCFTPLEALLSDLFR-DPDHCRQAYSVYAFMISLGGCLGY----------LLPAI 181
           |    :   |  |||:||   |    | | : |:| :::|  |||          | |
Db     153 DVANNMLQGPCRALLADLCNGDTRRMRSANAFYSFFMAVGNILGYAAGSYNNLYKLFPFS 212

Qy     182 DWDTSALAPYLGTQEEC-LFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPH 240
             | |   |    : | :  :: |||:| :| |  |  |    | |  |   :    |
Db     213 --KTHACDLYCANLKSCFIISIVLLIFITVLALT--VVREKQWSPDEADEEPPSSGKIP- 267

Qy     241 CCPCRARLAFRNL-GALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEG 299
                 |  | |||          :|| :  | |    :|:|   | || ||::|
Db     268 --------VFGELFGALKD--------LPRPMLMLLVVTCLNWIAWFPFILFDTDWMGRE 311

Qy     300 LYQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASV 359
           :| |   |  |    : ||:|||:|||| |   :  : |: :: ||:  |   : | :
Db     312 IYGGT--AGQG----KLYDQGVRVGSLGLLLNSVVLGLTSIAVEYLVRGVGGVKI-LWGL 364

Qy     360 AAFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTG 419
             | :| |      : || :  |                    |: :
Db     365 VNFLLAIGLV-----MTVVVSKVA------------------QHQRQH----------- 389

Qy     420 GASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGEPTE 479
                                   | :| | || |  |  |  :|:  ::| |
Db     390 ----------------------------GANGQLLPPSA--GVKAGALSLFSILGIPLS 418

Qy     480 ARV-VP----------GRGICLDLAILDSAFLLSQVAPSLFMGSIVQL--SQSVTAYMV 525
             :|           | |  | | :|: | :: |:  |:  |    |     :: |::|
Db     419 ITFSIPFALASIYSSGSGAGQGLSLGVLNLAIVVPQMIVSVLAGPFDSLFGGGNLPAFVV 478

Qy     526 ---SAAGLGLVAI 535
              |||  |::||
Db     479 GAISAAISGVLAI 491


RESULT  12
O04077
ID   O04077      PRELIMINARY;      PRT;   523 AA.
AC   O04077;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   SUCROSE TRANSPORT PROTEIN.
GN   SUT.
OS   Vicia faba (Broad bean).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids I; Fabales; Fabaceae; Papilionoideae; Vicieae; Vicia.
OX   NCBI_TaxID=3906;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=COTYLEDON;
RA   Weber H., Borisjuk L., Heim U., Sauer N., Wobus U.;
RT   "A role for sugar transporters during seed development: Molecular
RT   characterization of a hexose and a sucrose carrier in faba bean
RT   seeds.";
RL   Plant Cell 0:0-0(0).
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.
DR   EMBL; Z93774; CAB07811.1; -.
DR   InterPro; IPR003662; sub_transporter.
DR   Pfam; PF00083; sugar_tr; 1.
KW   Transmembrane.
SQ   SEQUENCE   523 AA;  55229 MW;  F8EBF170212D191A CRC64;


  Query Match             11.8%;  Score 337.5;  DB 10;  Length 523;
  Best Local Similarity   25.8%;  Pred. No. 4.6e-18;
  Matches  134;  Conservative   76;  Mismatches  185;  Indels  125;  Gaps   20;

Qy      17 QLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIGPVLGLVCVPLLGSASD 76
             ::::|  :  |::   |  :: : | :  :|:   :   :   ||: |::  |::|  ||
Db      36 KIMVVASIAAGVQFGWALQLSLLTPYVQLLGIHHTWAAYIWLCGPISGMLVQPIVGYHSD 95

Qy      77 HWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGL-------LCPDPRPLELALLILGV 129
               |:|||||| | |: : :::|||   |: | |        |    ||  : : ::|
Db      96 RCTSRFGRRRPFIAAGSIAVAIAVFLI---GYAADLGHSFGDSLDQKVRPRAIGIFVVGF 152

Qy     130 GLLDFCGQVCFTPLEALLSDLFR-DPDHCRQAYSVYAFMISLGGCLGYLLPAID------ 182
             :||    :   |  ||| ||   :   | | : ::| :::|  |||   |
Db     153 WILDVANNMLQGPCRALLGDLCAGNQRKTRNANAFFSFFMAVGNVLGYAAGAYSKLYHVF 212

Qy     183 --WDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLV-AEEAALGPTEPA----EGLSAP 235
               | |   |    : | |  |::  || :| : |:  :| || | :      :| |:
Db     213 PFTKTKACNVYCANLKSCFF--LSIALLTVLATSALIYVKETALTPEKTVVTTEDGGSSG 270

Qy     236 SLSPHCCPCRARL--AFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYT 293
           :     ||  :|  ||: |                | :  | :    :|:|   | || |
Db     271 GM-----PCFGQLSGAFKEL-------------KRPMWILLLVTCLNWIAWFPFLLFDT 311

Qy     294 DFVGEGLYQGVPRAEPGTEARRH-YDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTR 352
           |::|: :|       ||    | || ||| |:||| |   :    || :| | :  |
Db     312 DWMGKEVY-------GGTVGEGHAYDMGVREGALGLMLNSVVLGATSLGVDILARGVG-G 363

Qy     353 AVYLASVAAFPVAAGATCLSHSVAVV-------------------------TASAALTGF 387
             |  :  | :   | ||  :| |                          : ||| |
Db     364 VKRLWGIVNFLL---AICLGLTVLVTKLAQHSRQYAPGTGALGDPLPPSEGIKAGALTLF 420

Qy     388 TFSALQI-----LPYTLASLYHREKQVFLPKYRGDTGGASSEDSL-----------MTS 430
           :   : :     :|: |||::              | ||    ||             |
Db     421 SVLGVPLAITYSIPFALASIF-----------SSTSGAGQGLSLGVLNLAIVIPQMFVS 468

Qy     431 FLPGPKPG-------APFPNGHVGAGGSG-----LLPPPP 458
             | ||            |  | | |  ||     ||| ||
Db     469 VLSGPWDALFGGGNLPAFVVGAVAALASGILSIILLPSPP 508


RESULT  13
Q9SP63
ID   Q9SP63      PRELIMINARY;      PRT;   501 AA.
AC   Q9SP63;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   SUCROSE TRANSPORTER.
OS   Vitis vinifera (Grape).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Vitaceae;
OC   Vitis.
OX   NCBI_TaxID=29760;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CV. UGNI BLANC; TISSUE=GRAPE BERRY;
RA   Ageorges A., Issaly N., Picaud S., Delrot S., Romieu C.;
```

```
RT   "Identification and functional expression in yeast of a grape berry
RT   sucrose carrier.";
RL   Plant Physiol. Biochem. 38:177-185(2000).
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.
DR   EMBL; AF182445; AAD55269.1; -.
DR   InterPro; IPR003662; sub_transporter.
DR   Pfam; PF00083; sugar_tr; 1.
KW   Transmembrane.
SQ   SEQUENCE   501 AA;  54075 MW;  4FD19DFAFE539077 CRC64;
```

```
  Query Match             11.8%;  Score 336.5;  DB 10;  Length 501;
  Best Local Similarity   26.9%;  Pred. No. 5.2e-18;
  Matches  152;  Conservative   82;  Mismatches  199;  Indels  133;  Gaps   22;

Qy      17 QLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIGPVLGLVCVPLLGSASD 76
           :|| |  :  |::   |  :: : | : |:|:   : :::   ||: ||:  ||:|  ||
Db      29 RLLRVASVACGIQFGWALQLSLLTPYVQELGIPHAWSSIIWLCGPLSGLLVQPLVGHLSD 88

Qy      77 HWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLL--CPDPRPLELALLILGVGLLDF 134
              |:|||||| | :  |:::: :|  :  : |||    | ||  :|  ::|  |||
Db      89 RCNSRFGRRRPFIVAGATSIVVAVLIIGFSTDIGGLLGDGADRRPRAVATFVVGFWLLDV 148

Qy     135 CGQVCFTPLEALLSDLFRDPDH--CRQAYSVYAFMISLGGCLGYLLPA------IDWDTS 186
           |   |  |||:||  : ||   | | : ::  |::|  ||:   :       | | ||
Db     149 ANNVTQGPCRALLADL-TEKDHRRTRVANAYFSLFIAVGNVLGFATGSYSGWFRIFWFTS 207

Qy     187 ALAPYLGTQEECL----FGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPHCC 242
             :       :|       || :||   :| |  ::  ||          ||: | | |
Db     208 TSS----CNADCANLKSAFLLDIIF---IAITTYISITAA-----QELPLSSSSRSTHIS 255

Qy     243 -----PCRARLAFRNLGALLPRLHQLCCRMPRTLRR--------LFVAELCSWMALMTFT 289
                 |: ||         | :|      |||          |||  | :|:    |
Db     256 EEMAESTHAQEAF---------LWELF----GTLRYLSGSIWIILFVTAL-TWIGWFPFL 301

Qy     290 LFYTDFVGEGLYQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRF 349
           || ||::|  :| | |        ::|: |||||:||| |   :  : |::|::| :::
Db     302 LFDTDWMGREIYGGKP------NEGQNYNTGVRMGALGLMLNSVVLGITSVLMEKLCRKW 355

Qy     350 GTRAVYLASVAAFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQV 409
           |   |:                 |  |    |  : ||    | : | :
Db     356 GAGFVW----------------------GLSNILMSLCFLLMLILS---AVVKHMD--- 386

Qy     410 FLPKYRGDTGGASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVS 469
                                  ||    |    |:| | |           ||   | :
Db     387 ---------------------FLGHDLP----PSGVVIA----------ALIVFSILGIP 411

Qy     470 VRVVVGEP---TEARVVP-GRGICLDLAILDSAFLLSQVAPSLFMGSIVQL-----SQSV 520
           : :    |      |:   | |  | : :|: | :: ||  ||  |   ||      | |:
Db     412 LAITYSVPYALISTRIESLGLGQGLSMGVLNLAIVIPQVIVSLGSGPWDQLFGGGNSPSL 471

Qy     521 TAYMVSAAGLGLVAIYFATQVVFDKS 546
             |:|   |||||    :   |||
Db     472 AVAAVAAFASGLVAILAIPRSSADKS 497
```

```
RESULT  14
Q9ZTB9
ID   Q9ZTB9      PRELIMINARY;      PRT;   512 AA.
AC   Q9ZTB9;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   SUCROSE TRANSPORTER.
GN   SUT1.
OS   Apium graveolens (Celery).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC   Asteridae; euasterids II; Apiales; Apiaceae; Apium.
OX   TaxID=4045;
RN   [1]
RP   SEQUENCE FROM N.A.
```

```
RC   TISSUE=LEAF;
RA   Noiraud N., Delrot S., Lemoine R.;
RL   Submitted (MAY-1998) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.
DR   EMBL; AF063400; AAC99332.1; -.
DR   InterPro; IPR003662; sub_transporter.
DR   Pfam; PF00083; sugar_tr; 1.
KW   Transmembrane.
SQ   SEQUENCE   512 AA;  54520 MW;  4D76A7854A7DF903 CRC64;
```

```
  Query Match             11.6%;  Score 333;  DB 10;  Length 512;
  Best Local Similarity   25.0%;  Pred. No. 1e-17;
  Matches  139;  Conservative   86;  Mismatches  198;  Indels  132;  Gaps   21;

Qy      17 QLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIGPVLGLVCVPLLGSASD 76
           :|:||  :  |::   |  :: : | :  :|:  |:   :   ||: |::  |::|  ||
Db      33 KLILVAAIAAGVQFGWALQLSLLTPYVQLLGIPHKWAAYIWLCGPISGMLVQPIVGYYSD 92

Qy      77 HWRGRYGRRRPFIWALSLGILLSLFLIPRA---GWLAGLLCPDP-----RPLELALLILG 128
              :  :|||||| : :  : :|: ||  |    |: ||    |       :|  : : ::|
Db      93 RCQSSFGRRRPFIASGAGCVAISVILIGFAADIGYKAG----DDMNKTLKPRAVTVFVIG 148

Qy     129 VGLLDFCGQVCFTPLEALLSDLFR-DPDHCRQAYSVYAFMISLGGCLG---------YLL 178
           :||    :   |  |||:||    |    | | ::| | :::|  ||           | |
Db     149 FWILDVANNMLQGPCRALLADLCNGDTRRMRSANALYRFFMAVGNILGNAAGSYNNLYKL 208

Qy     179 PAIDWDTSALAPYLGTQEEC-LFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSL 237
             | |   |    : | :  :: |||:| :| |  |  |    | |  |   :
Db     209 FPFS-KTHACDLYCANLKSCFIISIVLLIFITVLALT--VVREKQWSPDEADEEPPSSGK 265

Qy     238 SPHCCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVG 297
           |              | ||  |  |    || :  |      :|:|   | || ||::|
Db     266 IP------------VFGELLRALKDL----PRPMLMLLAVTCLNWIAWFPFILFDTDWMG 309

Qy     298 EGLYQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLA 357
           :| |   |  |    : ||:|||:||||| |   :  : |: :: ||:  |   : |
Db     310 REIYGGT--AGQG----KLYDQGVRVGSLGLLLNSVVLGLTSIAVEYLVRGVGGVKI-LW 362

Qy     358 SVAAFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGD 417
           :  | :| |       : || :  |                        |: :
Db     363 GLVNFLLAIGLV-----MTVVVSKVA------------------QHQRQH--------- 389

Qy     418 TGGASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGEP 477
                                       | :| | || |  |  |  :|:  ::| |
Db     390 -------------------------------GANGQLLPPSA--GVKAGALSLFSILGIP 416

Qy     478 TEARV-VP----------GRGICLDLAILDSAFLLSQVAPSLFMGSIVQL--SQSVTAY 523
                 :|          | |  | | :|: | :: |:  |:  |    |    :: |:
Db     417 LSITFSIPFALASIYSSGSGAGQGLSLGVLNLAIVVPQMIVSVLAGPFDSLFGGGNLPAF 476

Qy     524 MV---SAAGLGLVAI 535
           :|   |||  |::||
Db     477 VVGAISAAISGVLAI 491
```

```
RESULT  15
Q43653
ID   Q43653      PRELIMINARY;      PRT;   516 AA.
AC   Q43653;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   SUCROSE TRANSPORT PROTEIN.
GN   SUT1.
OS   Solanum tuberosum (Potato).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC   Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
OX   NCBI_TaxID=4113;
RN   [1]
RP   SEQUENCE FROM N.A.
```

```
RC   STRAIN=CV. DESIREE; TISSUE=LEAF;
RX   MEDLINE=94146554; PubMed=8312741;
RA   Riesmeier J., Willmitzer L., Frommer W.B.;
RT   "Potato sucrose transporter expression in minor veins indicates a role
RT   in phloem loading.";
RL   Plant Cell 5:1591-1598(1993).
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.
DR   EMBL; X69165; CAA48915.1; -.
DR   InterPro; IPR003662; sub_transporter.
DR   Pfam; PF00083; sugar_tr; 1.
KW   Transmembrane.
SQ   SEQUENCE   516 AA;  54831 MW;  4FD06C095E49A377 CRC64;


  Query Match             11.5%;  Score 330;  DB 10;  Length 516;
  Best Local Similarity   24.7%;  Pred. No. 1.7e-17;
  Matches  132;  Conservative   93;  Mismatches  197;  Indels  112;  Gaps   21;

Qy       4 RLWVSRLLRHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIGPVL 63
           :||          ::::|  :  |::   |  :: : | :  :|:  || : :   ||:
Db      29 KLW----------KIIVVASIAAGVQFGWALQLSLLTPYVQLLGIPHKFASFIWLCGPIS 78

Qy      64 GLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGL-------LCPD 116
             |::  |::|  ||:   |:|||||||| | :  :::::|||   |: | |       |
Db      79 GMIVQPVVGYYSDNCSSRFGRRRPFIAAGAALVMIAVFLI---GFAADLGHASGDTLGKG 135

Qy     117 PRPLELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPD-HCRQAYSVYAFMISLGGCLG 175
            :|  :|: ::|  :||    :   |  |||:||        | | : ::| :::|  ||
Db     136 FKPRAIAVFVVGFWILDVANNMLQGPCRALLADLSGGKSGRMRTANAFFSFFMAVGNILG 195

Qy     176 YLLPAIDW--------DTSALAPYLGTQEECLF-GLLTLIFLTCVAATLLVAEE-AALGP 225
           |   :            | |   |    : | |  :  |: || :| ||:   |
Db     196 YAAGSYSHLFKVFPFSKTKACDMYCANLKSCFFIAIFLLLSLTTIALTLVRENELPEKDE 255

Qy     226 TEPAEGLSAPSLSPHCCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMAL 285
           | | |:           :::: |   | :   | :|    || :  | :    :|:|
Db     256 QEIDEKLAGAG--------KSKVPF--FGEIFGALKEL----PRPMWILLLVTCLNWIAW 301

Qy     286 MTFTLFYTDFVGEGLYQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRL 345
             | |: ||:: : :: |    : |    | || ||| |::|| ||  :    || :: |
Db     302 FPFFLYDTDWMAKEVFGG----QVGD--ARLYDLGVRAGAMGLLLQSVVLGFMSLGVEFL 355

Qy     346 VQRFGTRAVYLASVAAFPVAAGATCLSHSVAVV---------TASAALTGFT-------- 388
           :: |  |  |  :  | :   | ||: :: |             :  | | |
Db     356 GKKIG-GAKRLWGILNFVL---AICLAMTILVTKMAEKSRQHDPAGTLMGPTPGVKIGAL 411

Qy     389 --FSALQI-------LPYTLASLYHREK----------QVFLPKYRGDTGGASSEDS 426
             |:|| |       :|: |||::   :             : :|:      |   :|
Db     412 LLFAALGIPLAATFSIPFALASIFSSNRGSGQGLSLGVLNLAIVVPQMLVSLVGGPWDDL 471

Qy     427 LMTSFLPGPKPGAPFPNGHVGAGGSG-----LLPPPPALCGASACDVSVRVVVG 475
             |||       |  | | |  |      :|| |||       |    | :|
Db     472 FGGGNLPG------FVVGAVAAAASAVLALTMLPSPPA-------DAKPAVAMG 512


Search completed: June 26, 2002, 13:47:28
Job time: 611 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: June 26, 2002, 13:41:37 ; Search time 37.41 Seconds
(without alignments)
572.358 Million cell updates/sec

Title: US-09-838-785-2
Perfect score: 2861
Sequence: 1 MVQRLWVSRLLRHRKAQLLL.....AIYFATQVVFDKSDLAKYSA 553

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	489.5	17.1	530	1	MATP_MOUSE	P58355 mus musculu
2	488.5	17.1	530	1	MATP_HUMAN	Q9umx9 homo sapien
3	324.5	11.3	525	1	STP_SPIOL	Q03411 spinacia ol
4	142	5.0	544	1	YD74_SYNY3	P74168 synechocyst
5	136	4.8	541	1	GT10_HUMAN	O95528 homo sapien
6	125.5	4.4	399	1	TCR1_ECOLI	P02982 escherichia
7	121.5	4.2	503	1	PUR8_STRLP	P42670 streptomyce
8	119.5	4.2	491	1	AMPG_ECOLI	P36670 escherichia
9	117	4.1	440	1	YHJE_ECOLI	P37643 escherichia
10	117	4.1	495	1	SMVA_SALTY	P37594 salmonella
11	111.5	3.9	640	1	Y051_MYCTU	Q10880 mycobacteri
12	111	3.9	473	1	PHDK_NOCSK	O24723 nocardioide
13	110.5	3.9	680	1	CA1A_HUMAN	Q03692 homo sapien
14	110	3.8	368	1	GALT_HUMAN	O60755 homo sapien
15	110	3.8	606	1	NQOC_THETH	Q56227 thermus aqu
16	109.5	3.8	618	1	SPH2_HUMAN	Q9nra0 homo sapien
17	109	3.8	477	1	YPUM_RHOCA	P26176 rhodobacter
18	107	3.7	419	1	CMLA_PSEAE	P32482 pseudomonas
19	105	3.7	448	1	YJ94_YEAST	P47159 saccharomyc
20	104	3.6	476	1	MELB_SALTY	P30878 salmonella
21	103	3.6	465	1	FTSW_MYCLE	Q50186 mycobacteri
22	102.5	3.6	461	1	PUCC_RHOCA	P23462 rhodobacter
23	102	3.6	438	1	SHIA_ECOLI	P76350 escherichia
24	101.5	3.5	428	1	YXIO_BACSU	P42306 bacillus su
25	101	3.5	437	1	BRAZ_PSEAE	P25185 pseudomonas
26	101	3.5	471	1	MELB_ENTAE	O07366 enterobacte
27	101	3.5	471	1	MELB_KLEPN	Q02581 klebsiella
28	100	3.5	404	1	Y4XM_RHISN	P55705 rhizobium s
29	100	3.5	473	1	YIHO_SALTY	Q917r4 salmonella
30	99	3.5	481	1	LMRA_STRLN	P46104 streptomyce
31	98.5	3.4	613	1	NUOL_ECOLI	P33607 escherichia
32	98.5	3.4	685	1	FHUB_SALTY	O87656 salmonella
33	98	3.4	379	1	CYB_SORCI	O79451 sorex ciner
34	98	3.4	496	1	GTR3_CHICK	P28568 gallus gall
35	97.5	3.4	354	1	RNFD_PSEST	Q9evn4 pseudomonas
36	97.5	3.4	385	1	PI2R_BOVIN	P79393 bos taurus
37	97.5	3.4	433	1	DCUA_WOLSU	O34245 wolinella s
38	97.5	3.4	1103	1	CYGD_HUMAN	Q02846 homo sapien
39	97	3.4	336	1	CYB_SORGA	O79996 sorex grana
40	97	3.4	336	1	CYB_SORHA	O79452 sorex hayde
41	97	3.4	336	1	CYB_SORMO	O79969 sorex monti
42	97	3.4	336	1	CYB_SORVA	O80019 sorex vagra
43	97	3.4	370	1	GALT_RAT	O88626 rattus norv
44	97	3.4	427	1	RBTT_KLEPN	O52717 klebsiella
45	96.5	3.4	400	1	TCR8_PASMU	P51564 pasteurella

ALIGNMENTS

RESULT 1
MATP_MOUSE
ID MATP_MOUSE STANDARD; PRT; 530 AA.
AC P58355;
DT 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Membrane-associated transporter protein (AIM-1 protein) (Melanoma
DE antigen AIM1) (Underwhite protein).
GN MATP OR AIM1 OR UW.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DDY; TISSUE=Eye, Kidney, and Uterus;
RX MEDLINE=21372467; PubMed=11479596;
RA Fukamachi S., Shimada A., Shima A.;
RT "Mutations in the gene encoding B, a novel transporter protein, reduce
RT melanin content in medaka.";
RL Nat. Genet. 28:381-385(2001).
RN [2]
RP SEQUENCE FROM N.A., AND VARIANTS UW-DBR ASN-153 AND PRO-435.
RX MEDLINE=21473748; PubMed=11574907;
RA Newton J.M., Cohen-Barak O., Hagiwara N., Gardner J.M., Davisson M.T.,
RA King R.A., Brilliant M.H.;
RT "Mutations in the human orthologue of the mouse underwhite gene (uw)
RT underlie a new form of oculocutaneous albinism, OCA4.";
RL Am. J. Hum. Genet. 69:981-988(2001).
CC -!- FUNCTION: Melanocyte differentiation antigen. May transport
CC substances required for melanin biosynthesis (By similarity).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein; melanosome (By
CC similarity).
CC -!- TISSUE SPECIFICITY: Melanocytes, eyes, kidney and uterus.
CC -!- DISEASE: Defects in MATP are the cause of the UW-dbr phenotype
CC that results in loss of nearly all pigmentation in the homozygous
CC state.
CC -!- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.
CC ---

CC ---
DR EMBL; AF360357; AAK81713.1; -.
KW Melanin biosynthesis; Transmembrane; Antigen; Glycoprotein; Vision;
KW Disease mutation; Albinism.
FT DOMAIN 1 45 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 46 66 1 (POTENTIAL).
FT DOMAIN 67 68 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 69 89 2 (POTENTIAL).
FT DOMAIN 90 105 CYTOPLASMIC (POTENTIAL).

```
FT   TRANSMEM     106    126       3 (POTENTIAL).
FT   DOMAIN       127    138       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     139    159       4 (POTENTIAL).
FT   DOMAIN       160    184       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     185    205       5 (POTENTIAL).
FT   DOMAIN       206    216       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     217    237       6 (POTENTIAL).
FT   DOMAIN       238    318       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     319    339       7 (POTENTIAL).
FT   DOMAIN       340    366       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     367    387       8 (POTENTIAL).
FT   DOMAIN       388    398       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     399    419       9 (POTENTIAL).
FT   DOMAIN       420    425       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     426    446       10 (POTENTIAL).
FT   DOMAIN       447    477       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     478    498       11 (POTENTIAL).
FT   DOMAIN       499    504       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     505    525       12 (POTENTIAL).
FT   DOMAIN       526    530       CYTOPLASMIC (POTENTIAL).
FT   CARBOHYD     356    356       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARIANT      153    153       D -> N (IN UW-DBR).
FT   VARIANT      435    435       S -> P (IN UW-DBR).
SQ   SEQUENCE   530 AA;  57961 MW;  F4EDEAD07916D9FC CRC64;


  Query Match             17.1%;  Score 489.5;  DB 1;  Length 530;
  Best Local Similarity   26.1%;  Pred. No. 1.3e-27;
  Matches  146;  Conservative   85;  Mismatches  228;  Indels  101;  Gaps   12;

Qy      17 QLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIGPVLGLVCVPLLGSASD 76
           :|:: ::  || | | |    || |:|| ||: :   :||  : |:|| :  |::|||||
Db      34 RLVMHSMAMFGREFCYAVEAAYVTPVLLSVGLPKSLYSMVWLLSPILGFLLQPVVGSASD 93

Qy      77 HWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDPRP---LELALLILGVGLLD 133
           | | |:|||||:|  |:: :|| : |      :   |  :||      ::: ::|| | |
Db      94 HCRARWGRRRPYILTLAIMMLLGMALYLNGDAVVSALVANPRQKLIWAISITMVGVVLFD 153

Qy     134 FCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPAIDWDTSALAPYLG 193
           |       |::| | |:    |   :    :|     || |||:| ||||    |   ||
Db     154 FSADFIDGPIKAYLFDVCSHQDK-EKGLHYHALFTGFGGALGYILGAIDWVHLDLGRLLG 212

Qy     194 TQEECLFGLLTLIFLTCVAATLLVAEEAAL--GPTEP-------AEGLSAPSLSPHCCPC 244
           |: : :|    |: : |    |    || |    |:|          |||  :  :
Db     213 TEFQVMFFFSALVLILCFITHLCSIPEAPLRDAATDPPSQQDPQGSSLSASGMHEY---- 268

Qy     245 RARLAFRNLGA----------------------LLPRLHQLCCRMPRTLRRLFVAELCSW 282
             :    :| ||                      :  | :    ||   | | |: |  |
Db     269 GSIEKVKNGGADTEQPVQEWKNKKPSGQSQRTMSMKSLLRALVNMPSHYRCLCVSHLIGW 328

Qy     283 MALMTFTLFYTDFVGEGLYQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVM 342
           | ::  ||:|||:|: :| | |     :     |: || :|  || :    | |:|
Db     329 TAFLSNMLFFTDFMGQIVYHGDPYGAHNSTEFLIYERGVEVGCWGLCINSVFSSVYSYFQ 388

Qy     343 DRLVQRFGTRAVYLASVAAFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASL 402
             :|   | : :|      | :  |   |  :|       :: |   | |  :|: | :
Db     389 KAMVSYIGLKGLYFMGYLLFGLGTGFIGLFPNVYSTLVLCSMFGVMSSTLYTVPFNLIAE 448

Qy     403 YHREKQVFLPKYRGDTGGASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCG 462
           ||||::    | :|                   || |   |     | |:     |
Db     449 YHREEE----KEKGQEA-----------------PGGPDNQGR----GKGV------DCA 477

Qy     463 ASACDVSVRVVVGEPTEARVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTA 522
           |  | | :         |::: | |                       :| :| :: ||
Db     478 ALTCMVQL---------AQILVGGG---------------------LGFLVNMAGSVVV 506

Qy     523 YMVSAAGLGLVAIYFATQVV 542
            :::|: : |:   |   |
Db     507 VVITASAVSLIGCCFVALFV 526


RESULT   2
MATP_HUMAN

ID   MATP_HUMAN     STANDARD;      PRT;   530 AA.
AC   Q9UMX9; Q9BTM3;
DT   01-MAR-2002 (Rel. 41, Created)
DT   01-MAR-2002 (Rel. 41, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Membrane-associated transporter protein (AIM-1 protein) (Melanoma
DE   antigen AIM1).
GN   MATP OR AIM1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM AIM-1A).
RC   TISSUE=Melanoma;
RX   MEDLINE=21115844; PubMed=11221837;
RA   Harada M., Li Y.F., El-Gamil M., Rosenberg S.A., Robbins P.F.;
RT   "Use of an in vitro immunoselected tumor line to identify shared
RT   melanoma antigens recognized by HLA-A*0201-restricted T cells.";
RL   Cancer Res. 61:1089-1094(2001).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORM AIM-1B).
RC   TISSUE=Skin;
RA   Strausberg R.;
RL   Submitted (FEB-2001) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   ALTERNATIVE SPLICING.
RA   Ferro S.;
RL   Unpublished observations (NOV-2001).
RN   [4]
RP   DISEASE, AND VARIANT LEU-374.
RX   MEDLINE=21473748; PubMed=11574907;
RA   Newton J.M., Cohen-Barak O., Hagiwara N., Gardner J.M., Davisson M.T.,
RA   King R.A., Brilliant M.H.;
RT   "Mutations in the human orthologue of the mouse underwhite gene (uw)
RT   underlie a new form of oculocutaneous albinism, OCA4.";
RL   Am. J. Hum. Genet. 69:981-988(2001).
CC   -!- FUNCTION: Melanocyte differentiation antigen. May transport
CC       substances required for melanin biosynthesis (By similarity).
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein; melanosome (By
CC       similarity).
CC   -!- ALTERNATIVE PRODUCTS: At least 3 isoforms; AIM-1a (shown here),
CC       AIM-1b and AIM-1c; are produced by alternative splicing.
CC   -!- TISSUE SPECIFICITY: Expressed in most melanoma cell lines and
CC       melanocytes.
CC   -!- DISEASE: Defects in MATP are the cause of oculocutaneous albinism
CC       type 4 (OCA4). OCA4 is characterized by hypopigmentation of skin,
CC       hair and eyes. It leads to reduced visual acuity.
CC   -!- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.
CC   -!- CAUTION: Ref.2 sequence differs from that shown due to a
CC       frameshift in position 188.
CC   -!- CAUTION: The described alternatively spliced isoforms are infered
CC       using information from ests.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF172849; AAD51812.1; -.
DR   EMBL; BC003597; AAH03597.1; ALT_FRAME.
DR   MIM; 606202; -.
KW   Melanin biosynthesis; Transmembrane; Antigen; Glycoprotein; Vision;
KW   Polymorphism; Albinism; Alternative splicing.
FT   DOMAIN         1     46       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM      47     67       1 (POTENTIAL).
FT   DOMAIN        68     68       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM      69     89       2 (POTENTIAL).
FT   DOMAIN        90    110       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     111    131       3 (POTENTIAL).
```

```
FT   DOMAIN      132    138       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    139    159       4 (POTENTIAL).
FT   DOMAIN      160    184       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    185    205       5 (POTENTIAL).
FT   DOMAIN      206    216       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    217    237       6 (POTENTIAL).
FT   DOMAIN      238    318       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    319    339       7 (POTENTIAL).
FT   DOMAIN      340    366       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    367    387       8 (POTENTIAL).
FT   DOMAIN      388    398       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    399    419       9 (POTENTIAL).
FT   DOMAIN      420    425       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    426    446       10 (POTENTIAL).
FT   DOMAIN      447    477       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    478    498       11 (POTENTIAL).
FT   DOMAIN      499    504       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    505    525       12 (POTENTIAL).
FT   DOMAIN      526    530       CYTOPLASMIC (POTENTIAL).
FT   CARBOHYD    356    356       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARSPLIC    129    187       MISSING (IN ISOFORM AIM-1C).
FT   VARSPLIC    188    295       MISSING (IN ISOFORM AIM-1B).
FT   VARSPLIC    386    406       YFQKVLVSYIGLKGLYFTGYL -> CKSFSLLRMSSKSFWS
FT                                STTWI (IN AIM-1B).
FT   VARSPLIC    407    530       MISSING (IN AIM-1B).
FT   VARIANT     374    374       F -> L.
FT                                /FTId=VAR_012162.
SQ   SEQUENCE   530 AA;  58301 MW;  F14A4BACAA8FF31B CRC64;
```

```
  Query Match             17.1%;  Score 488.5;  DB 1;  Length 530;
  Best Local Similarity   26.4%;  Pred. No. 1.5e-27;
  Matches  150;  Conservative   78;  Mismatches  233;  Indels  107;  Gaps   12;

Qy       8 SRLLRHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIGPVLGLVC 67
           |||: |  |         || | | |    || |:|| ||:     ::|  : |:|| :
Db      33 SRLIMHSMAM--------FGREFCYAVEAAYVTPVLLSVGLPSSLYSIVWFLSPILGFLL 84

Qy      68 VPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDPR---PLELAL 124
           |::|||||| | |:|||||:|  | : :|: : |      :   |  :||      :::
Db      85 QPVVGSASDHCRSRWGRRRPYILTLGVMMLVGMALYLNGATVVAALIANPRRKLVWAISV 144

Qy     125 LILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPAIDWD 184
           ::|| | ||       |::| | |:    |   :    :|     || ||||| ||||
Db     145 TMIGVVLFDFAADFIDGPIKAYLFDVCSHQDK-EKGLHYHALFTGFGGALGYLLGAIDWA 203

Qy     185 TSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPHCCPC 244
             |  |||: : :|    |:  |    |    || |  || |:|:  |  :|   |
Db     204 HLELGRLLGTEFQVMFFFSALVLTLCFTVHLCSISEAPL--TEVAKGI-PPQQTPQDPPL 260

Qy     245 RA------------------RLAFRNL----------GALLPRLHQLCCRMPRTLRRL 274
           :                  || :                |  | :    ||   | |
Db     261 SSDGMYEYGSIEKVKNGYVNPELAMQGAKNKNHAEQTRRAMTLKSLLRALVNMPPHYRYL 320

Qy     275 FVAELCSWMALMTFTLFYTDFVGEGLYQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAI 334
           :: |  | | ::  ||:|||:|: :|:| | :   :     |: || :|  |  :
Db     321 CISHLIGWTAFLSNMLFFTDFMGQIVYRGDPYSAHNSTEFLIYERGVEVGCWGFCINSVF 380

Qy     335 SLVFSLVMDRLVQRFGTRAVYLASVAAFPVAAGATCLSHSVAVVTASAALTGFTFSALQI 394
           | ::|     ||   | : :|      | :  |   |  :|       :| |   | |
Db     381 SSLYSYFQKVLVSYIGLKGLYFTGYLLFGLGTGFIGLFPNVYSTLVLCSLFGVMSSTLYT 440

Qy     395 LPYTLASLYHREKQVFLPKYRGDTGGASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLL 454
           :|: | : ||||::    | |                 ||  |
Db     441 VPFNLITEYHREEE----KERQQA-------------PGGDP---------------- 465

Qy     455 PPPPALCGASACDVSVRVVVGEPTEARVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIV 514
                 | |||             |:|  :| | |     |:|:     :| :|
Db     466 -----------DNSVR------------GKG--MDCATLTCMVQLAQILVGGGLGFLV 498

Qy     515 QLSQSVTAYMVSAAGLGLVAIYFATQVV 542
           : :|   :::|: : |:   |    |
Db     499 NTAGTVVVVVITASAVALIGCCFVALFV 526
```

```
RESULT   3
STP_SPIOL
ID   STP_SPIOL      STANDARD;      PRT;   525 AA.
AC   Q03411;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Sucrose transport protein (Sucrose permease) (Sucrose-proton
DE   symporter).
OS   Spinacia oleracea (Spinach).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC   Caryophyllidae; Caryophyllales; Chenopodiaceae; Spinacia.
OX   NCBI_TaxID=3562;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Leaf;
RX   MEDLINE=93099843; PubMed=1464305;
RA   Riesmeier J.W., Willmitzer L., Frommer W.B.;
RT   "Isolation and characterization of a sucrose carrier cDNA from
RT   spinach by functional expression in yeast.";
RL   EMBO J. 11:4705-4713(1992).
CC   -!- FUNCTION: RESPONSIBLE FOR THE TRANSPORT OF SUCROSE INTO THE
CC       CELL, WITH THE CONCOMITANT EXPORT OF A PROTON (SYMPORT SYSTEM).
CC       CAN ALSO TRANSPORT MALTOSE AT A LESSER RATE.
CC   -!- PATHWAY: SUCROSE METABOLIC PATHWAY.
CC   -!- SUBCELLULAR LOCATION: INNER MEMBRANE.
CC   -!- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X67125; CAA47604.1; -.
DR   PIR; S28052; S28052.
DR   InterPro; IPR003662; sub_transporter.
DR   Pfam; PF00083; sugar_tr; 1.
DR   PROSITE; PS00216; SUGAR_TRANSPORT_1; FALSE_NEG.
DR   PROSITE; PS00217; SUGAR_TRANSPORT_2; FALSE_NEG.
KW   Transmembrane; Transport; Sugar transport; Symport.
FT   DOMAIN        1     37       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     38     58       1 (POTENTIAL).
FT   TRANSMEM     72     92       2 (POTENTIAL).
FT   TRANSMEM    107    127       3 (POTENTIAL).
FT   TRANSMEM    145    165       4 (POTENTIAL).
FT   TRANSMEM    184    204       5 (POTENTIAL).
FT   TRANSMEM    230    250       6 (POTENTIAL).
FT   TRANSMEM    295    315       7 (POTENTIAL).
FT   TRANSMEM    338    358       8 (POTENTIAL).
FT   TRANSMEM    373    393       9 (POTENTIAL).
FT   TRANSMEM    422    442       10 (POTENTIAL).
FT   TRANSMEM    455    475       11 (POTENTIAL).
FT   TRANSMEM    488    508       12 (POTENTIAL).
FT   DOMAIN      509    525       CYTOPLASMIC (POTENTIAL).
SQ   SEQUENCE   525 AA;  54992 MW;  018347A4D2CC1CC6 CRC64;
```

```
  Query Match             11.3%;  Score 324.5;  DB 1;  Length 525;
  Best Local Similarity   23.2%;  Pred. No. 6.4e-16;
  Matches  129;  Conservative   91;  Mismatches  196;  Indels  141;  Gaps   20;

Qy      23 LLTFGLEVCLAAGITY--------VPPLLLEVGVEEKFMTMVLGIGPVLGLVCVPLLGSA 74
           |   ||   :|||: :         : | :  :|:   :   :   ||: |::  ||:|
Db      35 LKKLGLVASVAAGVQFGWALQLSLLTPYVQLLGIPHTWAAYIWLCGPISGMIVQPLVGYY 94

Qy      75 SDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLA--GLLCPDP-----RPLELALLIL 127
           ||    |:|||||| | :  : ::: ||   |: |  |    ||     :|  :|: ::
```

```
Db      95 SDRCTSRFGRRRPFIAAGAALVAVAVGLI---GFAADIGAASGDPTGNVAKPRAIAVFVV 151

Qy     128 GVGLLDFCGQVCFTPLEALLSDLFR-DPDHCRQAYSVYAFMISLGGCLGYLLPAID---- 182

Db     152 GFWILDVANNTLQGPCRALLADMAAGSQTKTRYANAFFSFFMALGNIGGYAAGSYSRLYT 211

Qy     183 ----WDTSALAPYLGTQEECLFGLLT-LIFLTCVAATLLVAEEAALGPTEPAEGLSAPSL 237

Db     212 VFPFTKTAACDVYCANLKSCFFISITLLIVLTILALSVVKERQITIDEIQEEEDLKNRNN 271

Qy     238 SPHCCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVG 297

Db     272 SSGC----ARLPF--FGQLIGALKDL----PKPMLILLLVTALNWIAWFPFLLFDTDWMG 321

Qy     298 EGLYQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFG--TRAVY 355

Db     322 KEVYGGT------VGEGKLYDQGVHAGALGLMINSVVLGVMSLSIEGLARMVGGAKRLWG 375

Qy     356 LASVAAFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYR 415

Db     376 IVNIIL------AVCLAMTV-LVTKSAE------------------HFRDSHHIM---- 405

Qy     416 GDTGGASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVG 475

Db     406 ---------------------------------GSAVPPPPPA--GVKGGALAIFAVLG 429

Qy     476 EPTEARV-VP-----------GRGICLDLAILDSAFLLSQV----------------AP 506

Db     430 IPLAITFSIPFALASIFSASSGSGQGLSLGVLNLAIVVPQMFVSVTSGPWDAMFGGGNLP 489

Qy     507 SLFMGSIVQLSQSVTAY 523

Db     490 AFVVGAVAATASAVLSF 506


RESULT   4
YD74_SYNY3
ID   YD74_SYNY3     STANDARD;      PRT;   544 AA.
AC   P74168;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical symporter SLL1374.
GN   SLL1374.
OS   Synechocystis sp. (strain PCC 6803).
OC   Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX   NCBI_TaxID=1148;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97061201; PubMed=8905231;
RA   Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA   Miyajima N., Hirosawa M., Sugiura M., Sasamoto S., Kimura T.,
RA   Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,
RA   Shimpo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
RA   Tabata S.;
RT   "Sequence analysis of the genome of the unicellular cyanobacterium
RT   Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT   entire genome and assignment of potential protein-coding regions.";
RL   DNA Res. 3:109-136(1996).
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC   -!- SIMILARITY: BELONGS TO THE SODIUM:GALACTOSIDE SYMPORTER FAMILY
CC       (SGF).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D90912; BAA18257.1; -.
DR   InterPro; IPR001927; Na_galact_symp.
DR   Pfam; PF01236; Na_Galacto_symp; 1.
DR   PROSITE; PS00872; NA_GALACTOSIDE_SYMP; FALSE_NEG.
KW   Hypothetical protein; Transport; Transmembrane; Symport;
KW   Complete proteome.
FT   TRANSMEM      31     51       POTENTIAL.
FT   TRANSMEM      52     72       POTENTIAL.
FT   TRANSMEM      84    104       POTENTIAL.
FT   TRANSMEM     116    136       POTENTIAL.
FT   TRANSMEM     162    182       POTENTIAL.
FT   TRANSMEM     191    211       POTENTIAL.
FT   TRANSMEM     230    250       POTENTIAL.
FT   TRANSMEM     257    277       POTENTIAL.
FT   TRANSMEM     318    338       POTENTIAL.
FT   TRANSMEM     356    376       POTENTIAL.
FT   TRANSMEM     383    403       POTENTIAL.
FT   TRANSMEM     407    427       POTENTIAL.
FT   TRANSMEM     450    470       POTENTIAL.
FT   TRANSMEM     501    521       POTENTIAL.
SQ   SEQUENCE   544 AA;  60064 MW;  C91D0EDFF32277EE CRC64;


  Query Match             5.0%;  Score 142;  DB 1;  Length 544;
  Best Local Similarity   22.3%;  Pred. No. 0.0058;
  Matches  105;  Conservative   60;  Mismatches  179;  Indels  126;  Gaps   20;

Qy      16 AQLLLVNLLTFGLEVCLAAG----------ITYVPPLLLEV-GVEEKFMTMVLGIGPVL 63

Db       2 SQSLSAEKLHFTTKLAYGAGDFGPAITANILVFYLLFFLTDVAGIPAALAGSVLMIGKIF 61

Qy      64 GLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDPRPLELA 123

Db      62 DAINDPIIGLLSDRTRSRWGRRLPWM----LGGMIPFALFYTAQWLIPHFSDDRLTNQWG 117

Qy     124 LLILGVGLLDFCGQVCFT----PLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLP 179

Db     118 LFIYYVAIA-MAFNLCYTTVNLPYTALTPELTQNYNE-RTRLNSFRFAFSIGGSILSLIL 175

Qy     180 AIDWDTSALAPYLGTQEECLFGLL-TLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLS 238

Db     176 YI-----LIAAGLPDRPQQQFGELGVMISVLSISALLWSALRLQEKGKEP---ILSPSL- 226

Qy     239 PHCCPCRARLA-----------------FRNLGAL------------------------ 256

Db     227 ------RRRLAPLLMAAGITLILLAIAKSFNLLGGSGFDYISFFLILLGLIWGGFGFTLR 280

Qy     257 ----------------------LPRLHQLCCRMPRTLRR-LFV--AELCSWMALMTFTLF 291

Db     281 DSAVEEHLQKLENSPSPGVTENLPLLKQL--KIAFSNRAFLFVIGIYLCSWLAVQLTASI 338

Qy     292 YTDFVGEGLYQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGT 351

Db     339 LVYFVVS--WMGLNEQQSGT-------------IALAVQ-GTALVMLFVWQALAQFLDK 381

Qy     352 RAVY-LASVAAFPVAAGATCLS-HSVAVVTASAALTGFTFSALQILPYTL 399

Db     382 KVIYFLGSMVWMGAEAGLWLVQPGQVALLYTLAIFAGVGVSVAYLIPWSM 431


RESULT   5
GT10_HUMAN
ID   GT10_HUMAN     STANDARD;      PRT;   541 AA.
AC   O95528; Q9H4I6;
DT   01-MAR-2002 (Rel. 41, Created)
DT   01-MAR-2002 (Rel. 41, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Solute carrier family 2, facilitated glucose transporter, member 10
DE   (Glucose transporter type 10).
GN   SLC2A10 OR GLUT10.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
```

```
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=21145593; PubMed=11247674;
RA   McVie-Wylie A.J., Lamson D.R., Chen Y.T.;
RT   "Molecular cloning of a novel member of the GLUT family of
RT   transporters, SLC2A10 (GLUT10), localized on chromosome 20q13.1: a
RT   candidate gene for NIDDM susceptibility.";
RL   Genomics 72:113-117(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Stavrides G.S., Hashim Y., Huckle E.J., Deloukas P.;
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Fossey S.C., Mihic S.J., Craddock A.L., Mychaleckyj J.C., Dawson P.A.,
RA   Bowden D.W.;
RT   "GLUT10: a novel glucose transporter in the type 2 diabetes linked
RT   region of chromosome 20q12-13.1.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,
RA   Jones M., Stavrides G., Almeida J.P., Babbage A.K., Bagguley C.L.,
RA   Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M.,
RA   Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,
RA   Buck D., Burrill W., Butler A.P., Carder C., Carter N.P.,
RA   Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,
RA   Clegg S., Cobley V.E., Collier R.E., Connor R., Corby N.R.,
RA   Coulson A., Coville G.J., Deadman R., Dhami P., Dunn M.,
RA   Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,
RA   Grafham D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,
RA   Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,
RA   Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,
RA   Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,
RA   Lehvaslaiho M.H., Leversha M., Lloyd C., Lloyd D.M., Lovell J.D.,
RA   Marsh V.L., Martin S.L., McConnachie L.J., McLay K., McMurray A.A.,
RA   Milne S., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,
RA   Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,
RA   Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,
RA   Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Shownkeen R., Sims S.,
RA   Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,
RA   Swann M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,
RA   Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M.,
RA   Whitehead S.L., Whittaker P., Willey D.L., Williams L., Williams S.A.,
RA   Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,
RA   Rogers J.;
RT   "The DNA sequence and comparative analysis of human chromosome 20.";
RL   Nature 414:865-871(2001).
CC   -!- FUNCTION: Facilitative glucose transporter (By similarity).
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- TISSUE SPECIFICITY: Widely expressed; highest levels in liver and
CC       pancreas.
CC   -!- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY. GLUCOSE
CC       TRANSPORTERS SUBFAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF321240; AAK26294.1; -.
DR   EMBL; AL137188; CAB69822.2; -.
DR   EMBL; AF248053; AAK31911.1; -.
DR   EMBL; AL031055; CAA19926.2; -.
DR   MIM; 606145; -.
DR   InterPro; IPR003662; sub_transporter.
DR   InterPro; IPR003663; Sugar_transporter.
DR   Pfam; PF00083; sugar_tr; 1.
DR   PRINTS; PR00171; SUGRTRNSPORT.
DR   PROSITE; PS00216; SUGAR_TRANSPORT_1; 2.
DR   PROSITE; PS00217; SUGAR_TRANSPORT_2; FALSE_NEG.
KW   Transport; Sugar transport; Transmembrane; Glycoprotein;
KW   Multigene family.
FT   DOMAIN       1     15       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     16    36       1 (POTENTIAL).
FT   DOMAIN       37    48       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     49    69       2 (POTENTIAL).
FT   DOMAIN       70    77       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     78    98       3 (POTENTIAL).
FT   DOMAIN       99    106      EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     107   127      4 (POTENTIAL).
FT   DOMAIN       128   133      CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     135   155      5 (POTENTIAL).
FT   DOMAIN       156   166      EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     167   187      6 (POTENTIAL).
FT   DOMAIN       188   233      CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     234   254      7 (POTENTIAL).
FT   DOMAIN       255   269      EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     270   290      8 (POTENTIAL).
FT   DOMAIN       291   298      CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     299   319      9 (POTENTIAL).
FT   DOMAIN       320   414      EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     415   435      10 (POTENTIAL).
FT   DOMAIN       436   445      CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     446   466      11 (POTENTIAL).
FT   DOMAIN       467   476      EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     477   497      12 (POTENTIAL).
FT   DOMAIN       498   541      CYTOPLASMIC (POTENTIAL).
FT   CARBOHYD     334   334      N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   541 AA;  56911 MW;  6D644525FA136908 CRC64;

  Query Match             4.8%;  Score 136;  DB 1;  Length 541;
  Best Local Similarity   21.9%;  Pred. No. 0.015;
  Matches  130;  Conservative   72;  Mismatches  210;  Indels  182;  Gaps   27;

Qy      16 AQLLLVNLLTFGLEVCLAAGITYVPPLLLEVG---VEEKFMTMVLGIGPVLGLVCVPLLG 72
           |  : |:  |||| |: : :|   : || |: |   :|::|:   | :| :|  :    |
Db      12 ASVSLLGGLTFGYELAVISGA--LLPLQLDFGLSCLEQEFLVGSLLLGALLASLVGGFLI 69

Qy      73 SASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDPRPLELALLILGVGLL 132
                 |||::  : : :| :|     :  ||                 || |:||  ::
Db      70 DC-------YGRKQAILGS-NLVLLAGSLTLGLAG------------SLAWLVLGRAVV 108

Qy     133 DFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLG----GCLGYLLPAIDWDTS-- 186
           |   :        :|:|   |       |:|   |::|      | | |    |
Db     109 GFAISLSSMACCIYVSELV-GPRQRGVLVSLYEAGITVGILLSYALNYALAGTPWGWRHM 167

Qy     187 ---ALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPHCCP 243
              | ||        :   |:|:||   | |   | |      | :|  || | |
Db     168 FGWATAP-------AVLQSLSLLFLP--AGT---DETATHKDLIPLQGGEAPKLGPG--- 212

Qy     244 CRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGLYQG 303
           | | :| :|      :     |   |  :  :|     :: :    | |   | : |
Db     213 -RPRYSFLDLFRARDNMRG---RTTVGLGLVLFQQLTGQPNVLCYA--STIFSSVGFHGG 266

Qy     304 VPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVA--- 360
                         : |:||    |: :  :|    || | | ||: ||  |
Db     267 ---------------SSAVLASVGL---GAVKVAATLTAMGLVDRAGRRALLLAGCALMA 308

Qy     361 -----------AFPVAAGATCLSHSVAVVTASAALTG----FTFSALQILPYTLASLYHR 405
                     | |: :| :||  :|   |    | |        |:|  :| |
Db     309 LSVSGIGLVSFAVPMDSGPSCL--AVPNATGQTGLPGDSGLLQDSSLPPIPRT------N 360

Qy     406 EKQVFLPKYRGDTGGASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASA 465
           | |                |  | |:    | | :  |:       |  || ||
Db     361 EDQ--------------REPILSTAKKTKPHPRSGDPSAPPRLALSSALPGPP------- 399

Qy     466 CDVSVRVVVGEPTEARVVPGRG-------ICLDLAILDSAF----------LLSQVAP-- 506
                          :| ||         | | :  |||          :||:: |
Db     400 ----------------LPARGHALLRWTALLCLMVFVSAFSFGFGPVTWLVLSEIYPVE 442

Qy     507 ----------------SLFMG-------SIVQLSQSVTAYMVSAA-GLGLVAIY 536
```

```
                    :||:          : || :   | ::|  ||| : ::
Db    443 IRGRAFAFCNSFNWAANLFISLSFLDLIGTIGLSWTFLLYGLTAVLGLGFIYLF 496


RESULT   6
TCR1_ECOLI
ID   TCR1_ECOLI     STANDARD;      PRT;   399 AA.
AC   P02982;
DT   21-JUL-1986 (Rel. 01, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Tetracycline resistance protein, class A (TETA(A)).
GN   TETA.
OS   Escherichia coli.
OG   Plasmid RP1.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TRANSPOSON=Tn1721;
RX   MEDLINE=92192465; PubMed=1312499;
RA   Allmeier H., Cresnar B., Greck M., Schmitt R.;
RT   "Complete nucleotide sequence of Tn1721: gene organization and a
RT   novel gene product with features of a chemotaxis protein.";
RL   Gene 111:11-20(1992).
RN   [2]
RP   SEQUENCE OF 85-399 FROM N.A.
RC   STRAIN=DH1;
RA   Fletcher J.N., Hart C.A., Batt R.M., Saunders J.R.;
RL   Submitted (MAR-1994) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   PLASMID=RP1; TRANSPOSON=Tn1721;
RX   MEDLINE=83299270; PubMed=6310527;
RA   Waters S.H., Rogowsky P., Grinsted J., Altenbuchner J., Schmitt R.;
RT   "The tetracycline resistance determinants of RP1 and Tn1721:
RT   nucleotide sequence analysis.";
RL   Nucleic Acids Res. 11:6089-6105(1983).
RN   [4]
RP   TOPOLOGY.
RX   MEDLINE=92388137; PubMed=1517220;
RA   Allard J.D., Bertrand K.P.;
RT   "Membrane topology of the pBR322 tetracycline resistance protein.
RT   TetA-PhoA gene fusions and implications for the mechanism of TetA
RT   membrane insertion.";
RL   J. Biol. Chem. 267:17809-17819(1992).
CC   -!- FUNCTION: RESISTANCE TO TETRACYCLINE BY AN ACTIVE TETRACYCLINE
CC       EFFLUX. THIS IS AN ENERGY-DEPENDENT PROCESS THAT DECREASES THE
CC       ACCUMULATION OF THE ANTIBIOTIC IN WHOLE CELLS. THIS PROTEIN
CC       FUNCTIONS AS A METAL-TETRACYCLINE/H+ ANTIPORTER.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane.
CC   -!- SIMILARITY: BELONGS TO THE MAJOR FACILITATOR FAMILY (ALSO KNOWN
CC       AS THE DRUG RESISTANCE TRANSLOCASE FAMILY).
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X61367; CAA43643.1; -.
DR   EMBL; L29404; AAA83545.1; -.
DR   EMBL; X00006; CAA24909.1; -.
DR   PIR; A03509; YTECR1.
DR   InterPro; IPR001958; TCR_TetA.
DR   InterPro; IPR003662; sub_transporter.
DR   Pfam; PF00083; sugar_tr; 1.
DR   PRINTS; PR01035; TCRTETA.
DR   PROSITE; PS00216; SUGAR_TRANSPORT_1; UNKNOWN_1.
KW   Antibiotic resistance; Transmembrane; Inner membrane; Transport;
KW   Symport; Transposable element; Plasmid.
FT   DOMAIN        1      7       CYTOPLASMIC (PROBABLE).
FT   TRANSMEM      8     27       PROBABLE.
FT   DOMAIN       28     45       PERIPLASMIC (PROBABLE).
FT   TRANSMEM     46     66       PROBABLE.
FT   DOMAIN       67     79       CYTOPLASMIC (PROBABLE).
FT   TRANSMEM     80    100       PROBABLE.
FT   DOMAIN      101    103       PERIPLASMIC (PROBABLE).
FT   TRANSMEM    104    124       PROBABLE.
FT   DOMAIN      125    138       CYTOPLASMIC (PROBABLE).
FT   TRANSMEM    139    159       PROBABLE.
FT   DOMAIN      160    160       PERIPLASMIC (PROBABLE).
FT   TRANSMEM    161    181       PROBABLE.
FT   DOMAIN      182    210       CYTOPLASMIC (PROBABLE).
FT   TRANSMEM    211    231       PROBABLE.
FT   DOMAIN      232    246       PERIPLASMIC (PROBABLE).
FT   TRANSMEM    247    267       PROBABLE.
FT   DOMAIN      268    277       CYTOPLASMIC (PROBABLE).
FT   TRANSMEM    278    298       PROBABLE.
FT   DOMAIN      299    299       PERIPLASMIC (PROBABLE).
FT   TRANSMEM    300    320       PROBABLE.
FT   DOMAIN      321    339       CYTOPLASMIC (PROBABLE).
FT   TRANSMEM    340    360       PROBABLE.
FT   DOMAIN      361    364       PERIPLASMIC (PROBABLE).
FT   TRANSMEM    365    385       PROBABLE.
FT   DOMAIN      386    399       CYTOPLASMIC (PROBABLE).
FT   CONFLICT      5      5       R -> I (IN REF. 3).
FT   CONFLICT     55     55       M -> V (IN REF. 3).
FT   CONFLICT     75     75       V -> I (IN REF. 3).
FT   CONFLICT     84     84       A -> T (IN REF. 3).
FT   CONFLICT    201    203       ASF -> SFV (IN REF. 3).
SQ   SEQUENCE   399 AA;  42240 MW;  298427E6B5478374 CRC64;

  Query Match             4.4%;  Score 125.5;  DB 1;  Length 399;
  Best Local Similarity   22.3%;  Pred. No. 0.062;
  Matches  104;  Conservative   57;  Mismatches  136;  Indels  169;  Gaps   27;

Qy      19 LLVNLLTFGLEVCLAAGITYVPPL---LLEVGVEEKFMT----MVLGIGPVLGLVCVPLL 71
           |:| | |  |:   | ||  : |:   ||   |    :|    ::| :  ::   | |:|
Db       7 LIVILSTVALD---AVGIGLIMPVLPGLLRDLVHSNDVTAHYGILLALYALMQFACAPVL 63

Qy      72 GSASDHWRGRYGRR------------------RPFIWALSLGILLSLFLIPRAGWLAGL 112
           |: ||    |:|||                   ||:| | :| :::
Db      64 GALSD----RFGRRPVLLVSLAGAAVDYAIMATAPFLWVLYIGRIVA------------ 106

Qy     113 LCPDPRPLELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGG 172
                              |:    | |      | ::|:    :  |    : ||
Db     107 ----------------GITGATGAVA----GAYIADITDGDERARH----FGFM---SA 138

Qy     173 CLGYLLPAIDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGL 232
           | |: : |         | ||       ||:
Db     139 CFGFGMVA--------GPVLG-------GLM---------------------------- 154

Qy     233 SAPSLSPHCCPCRARLAFRNLG-----ALLPRLHQLCCRMPRTLRRLFVAELCS--W--- 282
               |||  |  |  |   |        |||  |:       | |||  :  | |  |
Db     155 --GGFSPH-APFFAAAALNGLNFLTGCFLLPESHK---GERRPLRREALNPLASFRWARG 208

Qy     283 ---MALMTFTLFYTDFVGEGLYQGVPRA--EPGTEARRHYDE---GVRMGSLGLFLQCAI 334
              :| :   |   ||:      || |      | | |:|    |: : : |:    |
Db     209 MTVVAALMAVFFIMQLVGQ-----VPAALWVIFGEDRFHWDATTIGISLAAFGILHSLAQ 263

Qy     335 SLVFSLVMDRLVQRFGTRAVYLASVA---AFPVAAGAT--CLSHSVAVVTASAALTGFTF 389
           :::   |  || :|   ||: |  :|    : : | ||   ::  : |: ||    |
Db     264 AMITGPVAARLGER---RALMLGMIADGTGYILLAFATRGWMAFPIMVLLASG---GIGM 317

Qy     390 SALQILPYTLASLYHREKQVFLPKYRGDTGGASSEDSLMTSFLPGP 435
             ||| :   |:     |:|       |   |: :  : :|| : ||
Db     318 PALQAM---LSRQVDEERQ-------GQLQGSLAALTSLTSIV-GP 352


RESULT   7
PUR8_STRLP
```

```
ID   PUR8_STRLP     STANDARD;      PRT;   503 AA.
AC   P42670;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   Puromycin resistance protein pur8.
GN   PUR8.
OS   Streptomyces lipmanii (Streptomyces alboniger).
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=1892;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 12461;
RX   MEDLINE=94109397; PubMed=7916693;
RA   Tercero J.A., Lacalle R.A., Jimenez A.;
RT   "The pur8 gene from the pur cluster of Streptomyces alboniger encodes
RT   a highly hydrophobic polypeptide which confers resistance to
RT   puromycin.";
RL   Eur. J. Biochem. 218:963-971(1993).
CC   -!- FUNCTION: MAY BE INVOLVED IN ACTIVE PUROMYCIN EFFLUX ENERGIZED BY
CC       A PROTON-DEPENDENT ELECTROCHEMICAL GRADIENT. IN ADDITION, IT COULD
CC       BE IMPLICATED IN SECRETING N-ACETYLPUROMYCIN, THE LAST
CC       INTERMEDIATE OF THE PUROMYCIN BIOSYNTHESIS PATHWAY, TO THE
CC       ENVIRONMENT.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- SIMILARITY: BELONGS TO THE MAJOR FACILITATOR FAMILY (ALSO KNOWN
CC       AS THE DRUG RESISTANCE TRANSLOCASE FAMILY).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X76855; CAA54186.1; -.
DR   Pfam; PF00083; sugar_tr; 1.
KW   Antibiotic resistance; Antibiotic biosynthesis; Transmembrane;
KW   Transport.
FT   DOMAIN        1     24       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     25     45       POTENTIAL.
FT   DOMAIN       46     64       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     65     85       POTENTIAL.
FT   DOMAIN       86     92       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     93    113       POTENTIAL.
FT   DOMAIN      114    122       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    123    143       POTENTIAL.
FT   DOMAIN      144    152       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    153    173       POTENTIAL.
FT   DOMAIN      174    181       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    182    202       POTENTIAL.
FT   DOMAIN      203    212       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    213    233       POTENTIAL.
FT   DOMAIN      234    241       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    242    262       POTENTIAL.
FT   DOMAIN      263    280       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    281    301       POTENTIAL.
FT   DOMAIN      302    313       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    314    334       POTENTIAL.
FT   DOMAIN      335    346       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    347    367       POTENTIAL.
FT   DOMAIN      368    371       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    372    392       POTENTIAL.
FT   DOMAIN      393    422       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    423    443       POTENTIAL.
FT   DOMAIN      444    461       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    462    482       POTENTIAL.
FT   DOMAIN      483    503       CYTOPLASMIC (POTENTIAL).
SQ   SEQUENCE   503 AA;  51852 MW;  FA3B814DF9A3EB74 CRC64;
```

```
  Query Match             4.2%;  Score 121.5;  DB 1;  Length 503;
  Best Local Similarity   23.2%;  Pred. No. 0.15;
  Matches  131;  Conservative   61;  Mismatches  175;  Indels  197;  Gaps   32;

Qy      26 FGLEVCLAAGITYV----------PPLLLEVGVEEKFMTMV-----LGIGPVLGLVCVPL 70
           :|| | ||| :  |          | :  ::|: :     |     |  | :|      |
Db      26 WGLVVILAAQLLVVLDGTVVNIALPSVQRDLGMSDTSRQWVITAYTLAFGGLL------L 79

Qy      71 LGSASDHWRGR----YGRRRPFIWALSLGILLSLFLIPRAGW-LAGLL---CPDPRPLEL 122
           ||       ||    :|||| |    ::|||          |: || ||    |||  | |
Db      80 LG-------GRVADAFGRRRIF----AVGIL---------GFGLASLLGGAAPDPGTLFL 119

Qy     123 ALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPAID 182
           |  : ||    |   :    | ||:: || :|    :|: ||  :   |  :| |
Db     120 ARALQGV----FAAALAPAAL-ALINTLFTEPGERGKAFGVYGAVSGGGAAVGLLAG--- 171

Qy     183 WDTSALAPYLGTQEECLFGLLT--LIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPH 240
                          ||||  | :  |:    ||  | ||
Db     172 -----------------GLLTEYLDWRWCLYVNAPVALLALLG---------------- 197

Qy     241 CCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAEL--CSWMALMTFTLFYTDFVGE 298
              ||          ||||      |  | :|      |  |  :  : :
Db     198 ---CR----------LLPRDR----RTGRAVRLDLPGTLLGCGGLVAIVYAF-------- 232

Qy     299 GLYQGVPRAEPGTEARRHYDEG--VRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYL 356
                       ||   : :   ||:  ||: :  | :||   | | |:
Db     233 -----------AEAESGWGDPLVVRLLVLGVLMLVAFALVERRVQDPLLP--------- 271

Qy     357 ASVAAFPVAAGATCLSHSVAVVTASAALTG----FTFSALQILPYT--LASLYHREKQVF 410
           | |  |  |:      : |     | |     |:     || |:  |  :       |
Db     272 PGVVAHRVRGGS-----FLVVGLPQIGLFGLFLFLTYYLQGILDYSPVLTGV------AF 320

Qy     411 LPKYRGDTGGASSEDSLMTSFLPGPKP-----GAPFPNGHVGAGGSGLL----PPPPALC 461
           ||   |   |:|    :    ||  :|     ||        | |  ||    |  |
Db     321 LPLGLGIAVGSS---LIAARLLPRTRPRTLIVGALL----AAAAGMALLTRLEPDTP--- 370

Qy     462 GASACDVSVRVVVGEPTEARVVPGRGI-CLDLAILDSAFLLSQVAP------SLFMGSIV 514
                 :| :     |::: | || |: :  : :|   ::|||      :  : |
Db     371 ---------QVYLTHLLPAQILIGLGIGCMMMPAMHTA--TARVAPHEAGAAAAVVNSAQ 419

Qy     515 QLSQSVTAYMVSAAGLGLVAIYFA 538
           |: ::   :::    |  | | |
Db     420 QVGGALGVALLNTVSTGATAAYLA 443
```

```
RESULT   8
AMPG_ECOLI
ID   AMPG_ECOLI     STANDARD;      PRT;   491 AA.
AC   P36670;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   AmpG protein.
GN   AMPG OR B0433 OR Z0536 OR ECS0487.
OS   Escherichia coli, and
OS   Escherichia coli O157:H7.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562, 83334;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94049112; PubMed=8231804;
RA   Lindquist S., Weston-Hafer K., Schmidt H., Pul C., Korfmann G.,
RA   Erickson J., Sanders C., Martin H.H., Normark S.;
RT   "AmpG, a signal transducer in chromosomal beta-lactamase induction.";
RL   Mol. Microbiol. 9:703-715(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SN0301-1, SN0301-3, AND SN0301-5;
RX   MEDLINE=95291453; PubMed=7773404;
RA   Schmidt H., Korfmann G., Barth H., Martin H.H.;
RT   "The signal transducer encoded by ampG is essential for induction of
RT   chromosomal AmpC beta-lactamase in Escherichia coli by beta-lactam
```

```
RT   antibiotics and 'unspecific' inducers.";
RL   Microbiology 141:1085-1092(1995).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE=97426617; PubMed=9278503;
RA   Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA   Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA   Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA   Mau B., Shao Y.;
RT   "The complete genome sequence of Escherichia coli K-12.";
RL   Science 277:1453-1474(1997).
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Roberts D., Allen E., Araujo R., Aparicio A., Chung E., Davis K.,
RA   Duncan M., Federspiel N., Hyman R., Kalman S., Komp C., Kurdi O.,
RA   Lew H., Lin D., Namath A., Oefner P., Schramm S., Davis R.W.;
RL   Submitted (JAN-1997) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE FROM N.A.
RC   STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX   MEDLINE=21074935; PubMed=11206551;
RA   Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA   Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA   Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA   Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA   Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA   Welch R.A., Blattner F.R.;
RT   "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL   Nature 409:529-533(2001).
RN   [6]
RP   SEQUENCE FROM N.A.
RC   STRAIN=O157:H7 / RIMD 0509952;
RX   MEDLINE=21156231; PubMed=11258796;
RA   Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA   Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA   Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA   Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT   "Complete genome sequence of enterohemorrhagic Escherichia coli
RT   O157:H7 and genomic comparison with a laboratory strain K-12.";
RL   DNA Res. 8:11-22(2001).
CC   -!- FUNCTION: PROBABLY ACTS AS A PERMEASE IN THE BETA-LACTAMASE
CC       INDUCTION SYSTEM AND IN PEPTIDOGLYCAN RECYCLING.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
CC       (Probable).
CC   -!- SIMILARITY: TO H.INFLUENZAE HI0350 AND TO YEAST YBR220C.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; S67816; AAB28884.1; -.
DR   EMBL; X82158; CAA57651.1; -.
DR   EMBL; X82159; CAA57652.1; -.
DR   EMBL; X82160; CAA57653.1; -.
DR   EMBL; AE000149; AAC73536.1; -.
DR   EMBL; U82664; AAB40189.1; -.
DR   EMBL; AE005222; AAG54783.1; -.
DR   EMBL; AP002551; BAB33910.1; -.
DR   PIR; S37391; S37391.
DR   EcoGene; EG12183; ampG.
DR   InterPro; IPR003662; sub_transporter.
DR   Pfam; PF00083; sugar_tr; 1.
KW   Transport; Transmembrane; Inner membrane; Complete proteome.
FT   TRANSMEM     13     33       POTENTIAL.
FT   TRANSMEM     81    105       POTENTIAL.
FT   TRANSMEM    110    128       POTENTIAL.
FT   TRANSMEM    174    197       POTENTIAL.
FT   TRANSMEM    224    244       POTENTIAL.
FT   TRANSMEM    266    283       POTENTIAL.
FT   TRANSMEM    328    345       POTENTIAL.
FT   TRANSMEM    382    406       POTENTIAL.
FT   TRANSMEM    425    450       POTENTIAL.
FT   TRANSMEM    461    485       POTENTIAL.
FT   VARIANT     151    151       G -> D (IN NON-INDUCIBLE MUTANT
FT                                SN0301-1).
FT   VARIANT     268    268       G -> D (IN NON-INDUCIBLE MUTANT
FT                                SN0301-3).
FT   VARIANT     373    373       G -> D (IN NON-INDUCIBLE MUTANT
FT                                SN0301-5).
SQ   SEQUENCE   491 AA;  53245 MW;  6B0D7E2BC699CA9C CRC64;


  Query Match             4.2%;  Score 119.5;  DB 1;  Length 491;
  Best Local Similarity   21.0%;  Pred. No. 0.2;
  Matches  126;  Conservative   76;  Mismatches  193;  Indels  205;  Gaps   30;

Qy       9 RLLRHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTM------VLGIGPV 62
           |: :  :: :||:         :  |:|:    || |  |  : :||:       :|   :
Db       7 RIFQQPRSAILLI--------LGFASGL----PLALTSGTLQAWMTVENIDLKTIGFFSL 54

Qy      63 LGLVCV------PLLGSASDHWRGRYGRRRPFIWALSLGILLSL----FLIP--RAGWLA 110
           :|   |      ||:   :  :   :|||| :: |  : :|:::    || |  :  |:|
Db      55 VGQAYVFKFLWSPLMDRYTPPF---FGRRRGWLLATQILLLVAIAAMGFLEPGTQLRWMA 111

Qy     111 GLLCPDPRPLELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDH-CRQAYSVYAF--- 166
           |                  :: ||        :|  :|:    :     | ||  :
Db     112 ALAV---------------VIAFCSASQDIVFDAWKTDVLPAEERGAGAAISVLGYRLG 155

Qy     167 MISLGGCLGYLLPAIDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPT 226
           |:  ||   :|  |  |        || |   :: |:  : : |: ||||  |     ||
Db     156 MLVSGGLALWL--ADKW--------LGWQG--MYWLMAALLIPCIIATLLAPE-----PT 198

Qy     227 EPAEGLSAPSLSPHCCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAEL------- 279
           :                          :|        :|:|| :  || |
Db     199 D--------------------------TIP--------VPKTLEQAVVAPLRDFFGRN 222

Qy     280 CSWMALMTFTL------FYTDFVGEGLYQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCA 333
           :|: |:   |      |        | :||     | :|    : ||   :|||
Db     223 NAWLILLLIVLYKLGDAFAMSLTTTFLIRGV-----GFDAG---EVGVVNKTLGLLATIV 274

Qy     334 ISLVFSLVMDRLVQRFGTRAVYLASVAAFPVAAGATCLSHSVAVVTASAALTGFTFSALQ 393
           :|   ::| ||      ||: :  :                          | | : :
Db     275 GALYGGILMQRLSL---FRALLIFGI----------------------LQGASNAGYW 307

Qy     394 ILPYTLASLYHREKQVFLPKYRGDTGGASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGL 453
           :|  |   ||      ||      |  | ::    |||           |   |   |
Db     308 LLSITDKHLYSMGAAVFFENLCGGMGTSAFVALLMT----------LCNKSFSATQFAL 356

Qy     454 LPPPPALCGASACDVSVRVVVGEPTEARVVPGRGICLDLAILDSAFLLSQVA---PSLFM 510
           |          ||      || || |      |   |          | | |  ||    | | :
Db     357 L---------SALSAVGRVYVG-PVAGWFVEAHGW--------STFYLFSVAAAVPGLIL 398

Qy     511 GSIVQLSQ----------SVTAYMVS--------AAGLGLVAIYFATQVVFDKSDLAKYS 552
           : : :              | |||           |||: |:|::    :  |  ||  :|
Db     399 LLVCRQTLEYTRVNDNFISRTAYPAGYAFAMWTLAAGVSLLAVWLLL-LTMDALDLTHFS 457


RESULT  9
YHJE_ECOLI
ID   YHJE_ECOLI     STANDARD;      PRT;   440 AA.
AC   P37643;
DT   01-OCT-1994 (Rel. 30, Created)
DT   01-OCT-1994 (Rel. 30, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Hypothetical metabolite transport protein yhjE.
GN   YHJE OR B3523.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
```

```
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE=94316500; PubMed=8041620;
RA   Sofia H.J., Burland V., Daniels D.L., Plunkett G. III, Blattner F.R.;
RT   "Analysis of the Escherichia coli genome. V. DNA sequence of the
RT   region from 76.0 to 81.5 minutes.";
RL   Nucleic Acids Res. 22:2576-2586(1994).
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
CC       (Potential).
CC   -!- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY. STRONG, TO
CC       H.INFLUENZAE HI0281 AND HI0418.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U00039; AAB18499.1; -.
DR   EMBL; AE000429; AAC76548.1; -.
DR   EcoGene; EG12249; yhjE.
DR   InterPro; IPR003662; sub_transporter.
DR   Pfam; PF00083; sugar_tr; 1.
DR   PROSITE; PS00216; SUGAR_TRANSPORT_1; 1.
DR   PROSITE; PS00217; SUGAR_TRANSPORT_2; FALSE_NEG.
KW   Hypothetical protein; Transport; Transmembrane; Inner membrane;
KW   Complete proteome.
FT   DOMAIN        1     34       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     35     55       1 (POTENTIAL).
FT   DOMAIN       56     66       PERIPLASMIC (POTENTIAL).
FT   TRANSMEM     67     87       2 (POTENTIAL).
FT   DOMAIN       88    108       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    109    129       3 (POTENTIAL).
FT   TRANSMEM    130    150       4 (POTENTIAL).
FT   DOMAIN      151    167       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    168    188       5 (POTENTIAL).
FT   DOMAIN      189    192       PERIPLASMIC (POTENTIAL).
FT   TRANSMEM    193    213       6 (POTENTIAL).
FT   DOMAIN      214    248       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    249    269       7 (POTENTIAL).
FT   DOMAIN      270    289       PERIPLASMIC (POTENTIAL).
FT   TRANSMEM    290    310       8 (POTENTIAL).
FT   DOMAIN      311    320       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    321    341       9 (POTENTIAL).
FT   DOMAIN      342    345       PERIPLASMIC (POTENTIAL).
FT   TRANSMEM    346    366       10 (POTENTIAL).
FT   DOMAIN      367    384       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    385    405       11 (POTENTIAL).
FT   DOMAIN      406    410       PERIPLASMIC (POTENTIAL).
FT   TRANSMEM    411    431       12 (POTENTIAL).
FT   DOMAIN      432    440       CYTOPLASMIC (POTENTIAL).
SQ   SEQUENCE   440 AA;  47208 MW;  3A23302A9DCBE5BE CRC64;


  Query Match             4.1%;  Score 117;  DB 1;  Length 440;
  Best Local Similarity   28.4%;  Pred. No. 0.27;
  Matches   48;  Conservative   29;  Mismatches   54;  Indels   38;  Gaps   10;

Qy      53 MTMVLGIGPVLGLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLI-PRAGWLAG 111
           |  |:| |     | ||: |  :|     :|||:  :   :| || :||   |  |
Db     295 MMAVIGFG-----VMVPVAGLLAD----AFGRRKSMVIITTLIILFALFAFNPLLG---- 341

Qy     112 LLCPDPRP-LELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISL 170
                 | |  | |:||: |:     : | |: ||| :||  |   |  |:  :|  ::
Db     342 ----SGNPILVFAFLLLGLSLMG----LTFGPMGALLPELF--PTEVR--YTGASFSYNV 389

Qy     171 GGCLGYLLPAIDWDTSALAPYLGTQEECLFGLLTL-IFLTCVAATLLVA 218
             ||          :::|||:    :  :||  : ::|  :|   |:|
Db     390 ASILG----------ASVAPYIAAWLQTNYGLGAVGLYLAAMAGLTLIA 428
```

```
RESULT  10
SMVA_SALTY
ID   SMVA_SALTY     STANDARD;      PRT;   495 AA.
AC   P37594;
DT   01-OCT-1994 (Rel. 30, Created)
DT   01-MAR-2002 (Rel. 41, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Methyl viologen resistance protein smvA.
GN   SMVA OR STM1574.
OS   Salmonella typhimurium.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Salmonella.
OX   NCBI_TaxID=602;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SL1303;
RX   MEDLINE=95011654; PubMed=7926834;
RA   Hongo E., Morimyo M., Mita K., Machida I., Hama-Inaba H., Tsuji H.,
RA   Ichimura S., Noda Y.;
RT   "The methyl viologen-resistance-encoding gene smvA of Salmonella
RT   typhimurium.";
RL   Gene 148:173-174(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX   MEDLINE=21534948; PubMed=11677609;
RA   McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA   Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA   Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA   Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA   Waterston R., Wilson R.K.;
RT   "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT   LT2.";
RL   Nature 413:852-856(2001).
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
CC       (Potential).
CC   -!- SIMILARITY: BELONGS TO THE MAJOR FACILITATOR FAMILY (ALSO KNOWN
CC       AS THE DRUG RESISTANCE TRANSLOCASE FAMILY).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D26057; BAA05055.1; -.
DR   EMBL; AE008769; AAL20492.1; -.
DR   StyGene; SG10384; smvA.
KW   Transmembrane; Inner membrane; Complete proteome.
FT   TRANSMEM      5     25       POTENTIAL.
FT   TRANSMEM     44     64       POTENTIAL.
FT   TRANSMEM     73     93       POTENTIAL.
FT   TRANSMEM     96    116       POTENTIAL.
FT   TRANSMEM    135    155       POTENTIAL.
FT   TRANSMEM    158    178       POTENTIAL.
FT   TRANSMEM    192    212       POTENTIAL.
FT   TRANSMEM    220    240       POTENTIAL.
FT   TRANSMEM    260    280       POTENTIAL.
FT   TRANSMEM    299    319       POTENTIAL.
FT   TRANSMEM    327    347       POTENTIAL.
FT   TRANSMEM    357    377       POTENTIAL.
FT   TRANSMEM    391    411       POTENTIAL.
FT   TRANSMEM    469    489       POTENTIAL.
FT   CONFLICT    144    144       A -> R (IN REF. 1).
FT   CONFLICT    147    147       G -> V (IN REF. 1).
FT   CONFLICT    182    182       V -> D (IN REF. 1).
FT   CONFLICT    198    198       A -> V (IN REF. 1).
FT   CONFLICT    227    227       F -> Y (IN REF. 1).
FT   CONFLICT    447    447       L -> Y (IN REF. 1).
FT   CONFLICT    454    457       AILD -> GKYLT (IN REF. 1).
SQ   SEQUENCE   495 AA;  52134 MW;  961F77C748CDA164 CRC64;
```

```
  Query Match             4.1%;  Score 117;  DB 1;  Length 495;
  Best Local Similarity   21.0%;  Pred. No. 0.31;
  Matches   92;  Conservative   60;  Mismatches  151;  Indels  136;  Gaps   22;

Qy      17 QLLLVNLLTFGL----------------EVCLAAGITYVPPLLLEVGV-----EEKFMT 54
           :||::    |||                | || |   : |  | |:     |||
Db      72 RLLMLGGTLFGLASLAAAFSHTASWLIATRVLLAIGAAMIVPATL-AGIRATFCEEKHRN 130

Qy      55 MVLGI-----------GPVLG-------------LVCVPL----LGSASDHWRGRYGRR- 85
           | ||:           ||::|             |: ||:    :|  : :   : |||
Db     131 MALGVWAAVGSGGAAFGPLIGGILLEHFYWGSVFLINVPIVLVVMGLTARYVPRQAGRRD 190

Qy      86 RPF---------------------------IWALSL-----GILLSLFLIPRAGWLAGL 112
           :|                           :| :|       :|| ||:  :      :
Db     191 QPLNLGHAVMLIIAILLLVYSAKTALKGHLSLWVISFTLLTGALLLGLFIRTQLATSRPM 250

Qy     113 LCPDPRPLELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMI---- 168
           :  | |     ::: || ::     :     | |::   :   |    |    ||:
Db     251 I--DMRLFTHRIILSGV-VMAMTAMITLVGFELLMAQELQFV-HGLSPYEAGVFMLPVMV 306

Qy     169 ------SLGGCL------------GYLLPAIDWDTSALAPYLGTQEECLFGLLTLIFLTC 210
                : | |             |  | |: :   |:  :  ||:   :||: |:  :
Db     307 ASGFSGPIAGVLVSRLGLRLVATGGMALSALSFYGLAMTDF-STQQWQAWGLMALLGFS- 364

Qy     211 VAATLLVAEEAALGPTEPAE-GLSAPSLSPHCCPCRARLAFRNLGALLPRLHQLCCRMPR 269
           ||: |:|  :|:    |||   :| ::        | |     | || |      |:|
Db     365 -AASALLASTSAIMAAAPAEKAAAAGAIETMAYELGAGLGIAIFGLLLSRSFSASIRLPA 423

Qy     270 TLRRLFVAELCSWMALMTFTLFYTDFVGEG--LYQGVP--RAEPGTEARRH---YDEGVR 322
           |    :|   | |             ||   |   :|  : :   :| ||   :     |
Db     424 GLEAQEIARASSSM------------GEAVQLANSLPPTQGQAILDAARHAFIWSHSVA 470

Qy     323 MGSLG-LFLQCAISLVFSL 340
           : | | : |  |: : |||
Db     471 LSSAGSMLLLLAVGMWFSL 489


RESULT  11
Y051_MYCTU
ID   Y051_MYCTU     STANDARD;      PRT;   640 AA.
AC   Q10880; O53628;
DT   01-OCT-1996 (Rel. 34, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical 65.6 kDa protein RV0083.
GN   RV0083 OR MT0090 OR MTCY251.01 OR MTV030.27.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37RV;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA   Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 1551 / Oshkosh;
RA   Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA   Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA   Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA   Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA   Bishai W.;
RT   "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT   laboratory strains.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC   -!- SIMILARITY: TO POLYPEPTIDE 5 OF THE NADH-UBIQUINOL OXIDOREDUCTASE
CC       OF CHLOROPLASTS OR MITOCHONDRIA.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AL021428; CAA16264.1; -.
DR   EMBL; AE006920; AAK44315.1; -.
DR   TIGR; MT0090; -.
DR   TubercuList; Rv0083; -.
DR   InterPro; IPR001750; Oxidored_q1.
DR   Pfam; PF00361; oxidored_q1; 1.
KW   Hypothetical protein; Oxidoreductase; Transmembrane;
KW   Complete proteome.
FT   TRANSMEM      8     28       POTENTIAL.
FT   TRANSMEM     52     72       POTENTIAL.
FT   TRANSMEM     90    110       POTENTIAL.
FT   TRANSMEM    136    156       POTENTIAL.
FT   TRANSMEM    179    199       POTENTIAL.
FT   TRANSMEM    208    228       POTENTIAL.
FT   TRANSMEM    241    261       POTENTIAL.
FT   TRANSMEM    277    297       POTENTIAL.
FT   TRANSMEM    298    318       POTENTIAL.
FT   TRANSMEM    352    372       POTENTIAL.
FT   TRANSMEM    391    411       POTENTIAL.
FT   TRANSMEM    446    466       POTENTIAL.
FT   TRANSMEM    497    517       POTENTIAL.
FT   TRANSMEM    619    639       POTENTIAL.
SQ   SEQUENCE   640 AA;  65627 MW;  6254C289DBD108A8 CRC64;


  Query Match             3.9%;  Score 111.5;  DB 1;  Length 640;
  Best Local Similarity   22.3%;  Pred. No. 0.99;
  Matches  129;  Conservative   73;  Mismatches  222;  Indels  155;  Gaps   29;

Qy      16 AQLLLV----NLLTFGL--EVCLAAGITYV-----PPLLLEVGVEEKFMTMVLGIGPVLG 64
           | :|||     :: || |  |:   | :  |      | :   |:    || :  |  ::|
Db      93 AAMLLVPAAGSVTTFLLAWELMAIASLILVLSEHARPQVRSAGLWYAVMTQLGFIAILVG 152

Qy      65 LVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLF-----------LIPRAGWLAGL 112
           || :   |  || : |       : |:  |:  ::|              |:|   ||
Db     153 LVVLAAAG-GSDRFAG--------LGAVCDGVRAAVFMLTLVGFGSKAGLVPLHAWLPRA 203

Qy     113 LCPDPRP----LELALLILGV-GLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFM 167
           | |    :  |:: ||: |:: |  |:           |   |     |      :
Db     204 HPEAPSPVSALMSAAMVNLGIYGIVRFDLQL----------LGPGPRWWGLA------L 246

Qy     168 ISLGG---CLGYLLPAIDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALG 224
           :::||     | |  ::  |   |  |  |:     ||:||     : |  | |:  | |
Db     247 LAVGGTSALYGVLQASVAADLKRLLAYSTTEN---MGLITL----ALGAATLFADTGAYG 299

Qy     225 PTEPAEGLSAPSLSPHCCPCRARLAFRNLGALLPR--------LHQLCCRMPRTLRRLFV 276
           |   |   :  : |       |||   |::|          |  |  ||| |
Db     300 PASIAAAAAMLHMIAHA--AFKSLAFMAAGSVLAATGLRDLDLLGGLARRMPAT------ 351

Qy     277 AELCSWMALMTFTLFY-------------TDFVGEG-LYQGVPRAEPGTEARRHYDEGVR 322
                   |:|:                || |  | | :  | || :        :
Db     352 ------------TVFFGVAALGACGLPLGAGFVSEWLLVQSLIHAAPGHDPIVALTTPLA 399

Qy     323 MGSLGLFLQCAISLVFS------LVMDRLVQRFGTRAVYLASVAAFPVAAGATCLSHSVA 376
           :| : |    ::: :         |   | |     |    :  |   :|||| ||  :||
Db     400 VGVVALATGLSVAAMTKAFGIGFLARPRSTQAEAAREAPASMRAGMAIAAGA-CLVLAVA 458
```

```
Qy     377 --VVTASAALTGFTFSALQILPYT-LASLYHREKQVFLPKYRGDTGGASSEDSLMTSFLP 433
           :|         |  | | : :| | ::      | ||   |                |  |
Db     459 PLLVAPMVRRAAATLPAAQAVKFTGLGAV------VRLPAMSG------------SIAP 499

Qy     434 GPKPGAPFPNGHVGAGGSGLLP-------PPPA-----LCGASACDVSVRVVVGEPTEAR 481
           |    |       |    :|       | ||      |||:  |::||:     : |
Db     500 GVIAAAVL----AAALAVAVLARWRFRRRPAPARLPLWACGAA--DLTVRMQYTATSFAE 553

Qy     482 VVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSV 520
           | : :  |:   |:   ::  | | :|   :    :|
Db     554 --PLQRVFGDVLRPDTDIEVTHTAESRYMAERITYRTAV 590
```

RESULT 12
PHDK_NOCSK
ID PHDK_NOCSK STANDARD; PRT; 473 AA.
AC O24723;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Probable 1-hydroxy-2-naphthoate transporter.
OS Nocardioides sp. (strain KP7).
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Propionibacterineae; Nocardioidaceae; Nocardioides.
OX NCBI_TaxID=35761;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97474276; PubMed=9335300;
RA Iwabuchi T., Harayama S.;
RT "Biochemical and genetic characterization of 2-carboxybenzaldehyde
RT dehydrogenase, an enzyme involved in phenanthrene degradation by
RT Nocardioides sp. strain KP7.";
RL J. Bacteriol. 179:6488-6494(1997).
CC -!- FUNCTION: PROBABLY INVOLVED IN THE TRANSPORT OF 1-HYDROXY-2-
CC NAPHTHOATE.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
CC (Potential).
CC -!- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.
CC ---

CC ---
DR EMBL; AB000735; BAA23264.1; -.
DR InterPro; IPR003662; sub_transporter.
DR Pfam; PF00083; sugar_tr; 1.
DR PROSITE; PS00216; SUGAR_TRANSPORT_1; 1.
DR PROSITE; PS00217; SUGAR_TRANSPORT_2; FALSE_NEG.
KW Transport; Transmembrane; Inner membrane.
FT DOMAIN 1 25 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 26 46 1 (POTENTIAL).
FT DOMAIN 47 59 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 60 80 2 (POTENTIAL).
FT DOMAIN 81 91 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 92 112 3 (POTENTIAL).
FT DOMAIN 113 121 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 122 142 4 (POTENTIAL).
FT DOMAIN 143 153 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 154 174 5 (POTENTIAL).
FT DOMAIN 175 180 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 181 201 6 (POTENTIAL).
FT DOMAIN 202 263 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 264 284 7 (POTENTIAL).
FT DOMAIN 285 304 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 305 325 8 (POTENTIAL).
FT DOMAIN 326 330 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 331 351 9 (POTENTIAL).
FT DOMAIN 352 354 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 355 375 10 (POTENTIAL).
FT DOMAIN 376 398 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 399 418 11 (POTENTIAL).
FT DOMAIN 419 421 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 422 444 12 (POTENTIAL).
FT DOMAIN 445 473 CYTOPLASMIC (POTENTIAL).
SQ SEQUENCE 473 AA; 49109 MW; D6D765D376260D8A CRC64;

Query Match 3.9%; Score 111; DB 1; Length 473;
Best Local Similarity 24.3%; Pred. No. 0.78;
Matches 114; Conservative 66; Mismatches 180; Indels 110; Gaps 27;

```
Qy      12 RHRKAQLLLVNLLTF--GLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIGPVLGLVCVP 69
           | ::  :|:| |:    |::: ||: :   ||:: : ||    :|:|: :| |   :
Db      22 RRQRIIVLMVFLIMVADGMDITLASHL--FPPVIRDWGVPVSAVTLVVSLGVVAMAIGAL 79

Qy      70 LLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAG-WLAGLLCPDPRPLELALLILG 128
           : |  :| | || |          :| :  || :  ||  | |    |         :|
Db      80 VSGPVADRW-GRKG-------VTVVGFV--LFCLATAGLGLTG----DIHSFAALRIISC 125

Qy     129 VGLLDFCGQVCFTPLEALLSDLFRDPDHCR-QAYSVYAFMISLGGCLGYLLPAIDWDTSA 187
            ||    | |    |  :::|    |   | |  |:    : :|  :|  | |      |
Db     126 FGL----GAVMPVAL-TIVADWM--PKARRAQMVSIAFAGVGVGSIIGAYLAA------A 172

Qy     188 LAPYLGTQEECLF-GLLTLIFLTCVAATLLVAEEA--------------ALGPTEPAEG 231
           : | || |   |  ||  || |    |  || | |                ||    |
Db     173 VIPTLGWQVMVLIAGLAPLIILPFFVA--LVPEPAIISVRRGIPEARIRSALALVAPDRD 230

Qy     232 LSAPSLSPHCCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLF 291
           ::   |:      || |    |||   |            : ||   ||  :  :|  ::
Db     231 IAGVDLT------RAGL---TLGAGEVR----------AKALFAEILCRPLLGVTLLIW 270

Qy     292 YTDFVGEG----LYQGVP----RAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMD 343
             || :|    : | :|       ||       : |: :   |       :|:  | :
Db     271 GVFFVVQGSGLLVLQYMPMLLQAPAPGLST---VESGLIVAMYGWG-----ALIGQLTIA 322

Qy     344 RLVQRFGTRAVYLASVAAFPVAAGATCLSHSVAVVTASAALTGFTF-------SALQI-L 395
           :::||  | : ||:   : |          | ::  :|  ||| |        |: : |
Db     323 FILKRF-DRFIALAAFIFWSV----------VGLLIVAAFGTGFGFFGYFTLLFAIGLSL 371

Qy     396 PYTLASLYHREKQVFLPKYR----GDTGGASSEDSLMTSFLPGPKPGAPF 441
           | | |::       :  ::|   '|  | |    :|    | |   || |
Db     372 PATAAAMQSVTTLAYEEEFRATGMGSAGFAGRLGTLTYGALGGTLIGAGF 421
```

RESULT 13
CA1A_HUMAN
ID CA1A_HUMAN STANDARD; PRT; 680 AA.
AC Q03692;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Collagen alpha 1(X) chain precursor.
GN COL10A1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92109659; PubMed=1764025;
RA Thomas J.T., Cresswell C.J., Rash B., Nicolai H., Jones T.,
RA Solomon E., Grant M.E., Boot-Handford R.P.;
RT "The human collagen X gene. Complete primary translated sequence and
RT chromosomal localization.";
RL Biochem. J. 280:617-623(1991).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=93012005; PubMed=1397333;
RA Reichenberger E., Beier F., Luvalle P., Olsen B.R., von der Mark K.,
RA Bertling W.M.;
RT "Genomic organization and full-length cDNA sequence of human collagen

```
RT   X.";
RL   FEBS Lett. 311:305-310(1992).
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Beier F., Lammi M.B., von der Mark K.;
RL   Submitted (JUN-1996) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Williams S.;
RL   Submitted (JAN-2001) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE OF 52-680 FROM N.A.
RX   MEDLINE=92267014; PubMed=1587271;
RA   Apte S.S., Seldin M.F., Hayashi M., Olsen B.R.;
RT   "Cloning of the human and mouse type X collagen genes and mapping of
RT   the mouse type X collagen gene to chromosome 10.";
RL   Eur. J. Biochem. 206:217-224(1992).
RN   [6]
RP   SEQUENCE OF 561-666 FROM N.A.
RX   MEDLINE=91243838; PubMed=2037056;
RA   Apte S., Mattei M.-G., Olsen B.R.;
RT   "Cloning of human alpha 1(X) collagen DNA and localization of the
RT   COL10A1 gene to the q21-q22 region of human chromosome 6.";
RL   FEBS Lett. 282:393-396(1991).
RN   [7]
RP   SEQUENCE OF 547-655 FROM N.A.
RX   MEDLINE=92077285; PubMed=1743401;
RA   Reichenberger E., Aigner T., von der Mark K., Stoeb H., Bertling W.;
RT   "In situ hybridization studies on the expression of type X collagen
RT   in fetal human cartilage.";
RL   Dev. Biol. 148:562-572(1991).
RN   [8]
RP   REVIEW ON VARIANTS.
RX   MEDLINE=97255959; PubMed=9101290;
RA   Kuivaniemi H., Tromp G., Prockop D.J.;
RT   "Mutations in fibrillar collagens (types I, II, III, and XI), fibril-
RT   associated collagen (type IX), and network-forming collagen (type X)
RT   cause a spectrum of diseases of bone, cartilage, and blood vessels.";
RL   Hum. Mutat. 9:300-315(1997).
RN   [9]
RP   VARIANTS SMCD ASP-598 AND PRO-614.
RX   MEDLINE=94136476; PubMed=8304336;
RA   Wallis G.A., Rash B., Sweetman W.A., Thomas J.T., Super M., Evans G.,
RA   Grant M.E., Boot-Handford R.P.;
RT   "Amino acid substitutions of conserved residues in the
RT   carboxyl-terminal domain of the alpha 1(X) chain of type X collagen
RT   occur in two unrelated families with metaphyseal chondrodysplasia
RT   type Schmid.";
RL   Am. J. Hum. Genet. 54:169-178(1994).
RN   [10]
RP   VARIANT SMCD ARG-591.
RX   MEDLINE=94272470; PubMed=8004099;
RA   McIntosh I., Abbott M.H., Warman M.L., Olsen B.R., Francomano C.A.;
RT   "Additional mutations of type X collagen confirm COL10A1 as the
RT   Schmid metaphyseal chondrodysplasia locus.";
RL   Hum. Mol. Genet. 3:303-307(1994).
RN   [11]
RP   VARIANT SMCD VAL-618.
RX   MEDLINE=95181449; PubMed=7876225;
RA   Chan D., Cole W.G., Rogers J.G., Bateman J.F.;
RT   "Type X collagen multimer assembly in vitro is prevented by a Gly618
RT   to Val mutation in the alpha 1(X) NC1 domain resulting in Schmid
RT   metaphyseal chondrodysplasia.";
RL   J. Biol. Chem. 270:4558-4562(1995).
RN   [12]
RP   VARIANTS SMCD ARG-545; GLU-595; HIS-597; LYS-617; ARG-644 AND GLY-648.
RX   MEDLINE=95331767; PubMed=7607655;
RA   Bonaventure J., Chaminade F., Maroteaux P.;
RT   "Mutations in three subdomains of the carboxy-terminal region of
RT   collagen type X account for most of the Schmid metaphyseal
RT   dysplasias.";
RL   Hum. Genet. 96:58-64(1995).
RN   [13]
RP   VARIANT SMCD PRO-600.
RX   MEDLINE=96375754; PubMed=8782043;
RA   Wallis G.A., Rash B., Sykes B., Bonaventure J., Maroteaux P.,
RA   Zabel B., Wynne-Davies R., Grant M.E., Boot-Handford R.P.;
RT   "Mutations within the gene encoding the alpha 1 (X) chain of type X
RT   collagen (COL10A1) cause metaphyseal chondrodysplasia type Schmid but
RT   not several other forms of metaphyseal chondrodysplasia.";
RL   J. Med. Genet. 33:450-457(1996).
RN   [14]
RP   VARIANTS SMCD GLU-18 AND ARG-18.
RX   MEDLINE=97220591; PubMed=9067753;
RA   Ikegawa S., Nakamura K., Nagano A., Haga N., Nakamura Y.;
RT   "Mutations in the N-terminal globular domain of the type X collagen
RT   gene (COL10A1) in patients with schmid metaphyseal
RT   chondrodysplasia.";
RL   Hum. Mutat. 9:131-135(1997).
RN   [15]
RP   VARIANTS SMD GLU-595.
RX   MEDLINE=99057503; PubMed=9837818;
RA   Ikegawa S., Nishimura G., Nagai T., Hasegawa T., Ohashi H.,
RA   Nakamura Y.;
RT   "Mutation of the type X collagen gene 'COL10A1' causes
RT   spondylometaphyseal dysplasia.";
RL   Am. J. Hum. Genet. 63:1659-1662(1998).
RN   [16]
RP   VARIANT SMCD CYS-597.
RX   MEDLINE=99069781; PubMed=9852679;
RA   Sawai H., Ida A., Nakata Y., Koyama K.;
RT   "Novel missense mutation resulting in the substitution of tyrosine by
RT   cysteine at codon 597 of the type X collagen gene associated with
RT   Schmid metaphyseal chondrodysplasia.";
RL   J. Hum. Genet. 43:259-261(1998).
CC   -!- FUNCTION: TYPE X COLLAGEN IS A PRODUCT OF HYPERTHROPHIC
CC       CHONDROTOCYTES AND HAS BEEN LOCALIZED TO PRESUMPTIVE
CC       MINERALIZATION ZONES OF HYALINE CARTILAGE.
CC   -!- SUBUNIT: HOMOTRIMER.
CC   -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC       UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC   -!- DISEASE: DEFECTS IN COL10A1 ARE THE CAUSE OF SCHMID TYPE
CC       METAPHYSEAL CHONDRODYSPLASIA (SMCD), A DOMINANTLY INHERITED
CC       DISORDER OF THE OSSEOUS SKELETON. THE CARDINAL FEATURES OF THE
CC       PHENOTYPE ARE MILD SHORT STATURE, COXA VARA AND A WADDLING GAIT.
CC       RADIOGRAPHY USUALLY SHOWS SCLEROSIS OF THE RIBS, FLARING OF THE
CC       METAPHYSES, AND A WIDE IRREGULAR GROWTH PLATE, ESPECIALLY OF THE
CC       KNEES.
CC   -!- DISEASE: DEFECTS IN COL10A1 ARE THE CAUSE OF SPONDYLOMETAPHYSEAL
CC       DYSPLASIA (SMD). SMD COMPRISES A HETEROGENEOUS GROUP OF HERITABLE
CC       SKELETAL DYSPLASIAS CHARACTERIZED BY MODIFICATIONS OF THE
CC       VERTEBRAL BODIES OF THE SPINE AND METAPHYSES OF THE TUBULAR BONES.
CC   -!- SIMILARITY: STRONG, TO ALPHA 1 AND 2 TYPE VIII COLLAGENS.
CC   -!- SIMILARITY: CONTAINS 1 C1Q DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X60382; CAA42933.1; -.
DR   EMBL; X65120; CAA46236.1; -.
DR   EMBL; X98568; CAA67178.1; -.
DR   EMBL; AL121963; CAB87590.1; -.
DR   EMBL; S68531; AAC60615.1; -.
DR   EMBL; X58879; CAA41686.1; -.
DR   EMBL; M74050; AAA61221.1; -.
DR   EMBL; X72579; CAA51170.1; -.
DR   EMBL; X72580; CAA51170.1; JOINED.
DR   PIR; S15826; S15826.
DR   PIR; S30086; S30086.
DR   PIR; A43901; A43901.
DR   PIR; S18249; S18249.
```

```
DR   PIR; S21856; S21856.
DR   PIR; S26396; S26396.
DR   MIM; 120110; -.
DR   MIM; 156500; -.
DR   MIM; 184250; -.
DR   InterPro; IPR001073; C1q.
DR   InterPro; IPR000087; Collagen.
DR   Pfam; PF00386; C1q; 1.
DR   Pfam; PF01391; Collagen; 7.
DR   PRINTS; PR00007; COMPLEMNTC1Q.
DR   SMART; SM00110; C1Q; 1.
DR   PROSITE; PS01113; C1Q; 1.
KW   Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW   Cartilage; Collagen; Signal; Disease mutation; Polymorphism.
FT   SIGNAL        1     18       POTENTIAL.
FT   CHAIN        19    680       COLLAGEN ALPHA 1(X) CHAIN.
FT   DOMAIN       19     56       NONHELICAL REGION (NC2).
FT   DOMAIN       57    519       TRIPLE-HELICAL REGION.
FT   DOMAIN      520    680       NONHELICAL REGION (NC1).
FT   DOMAIN      545    680       C1Q.
FT   VARIANT      18     18       G -> E (IN SMCD).
FT                                /FTId=VAR_001838.
FT   VARIANT      18     18       G -> R (IN SMCD).
FT                                /FTId=VAR_001839.
FT   VARIANT     545    545       G -> R.
FT                                /FTId=VAR_001840.
FT   VARIANT     591    591       C -> R (IN SMCD).
FT                                /FTId=VAR_001841.
FT   VARIANT     595    595       G -> E (IN SMCD AND SMD).

  Query Match             3.9%;  Score 110.5;  DB 1;  Length 680;
  Best Local Similarity   30.7%;  Pred. No. 1.2;
  Matches   43;  Conservative   11;  Mismatches   51;  Indels   35;  Gaps    7;

Qy     412 PKYRGDTG--GASSEDSLMTSFLPGP--KPGAPFPNGHVGAGGSGLLPPPPALCGASACD 467
           |  :|| |  |      : |  | ||   || | | || |  |   || ||
Db     463 PGSKGDPGSPGPPGPAGIATKGLNGPTGPPGPPGPRGHSGEPG---LPGPP--------- 510

Qy     468 VSVRVVVGEPTEARVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSA 527
               | |      ||:      |::   |: :   |||    :|  :| ||   |||
Db     511 -------GPPGP----PGQ------AVMPEGFIKAGQRPSLSGTPLVSANQGVTGMPVSA 553

Qy     528 AGLGLVAIY--FATQVVFDK 545
           : |   |    | : |||
Db     554 FTVILSKAYPAIGTPIPFDK 573


RESULT  14
GALT_HUMAN
ID   GALT_HUMAN     STANDARD;      PRT;   368 AA.
AC   O60755;
DT   15-JUL-1999 (Rel. 38, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Galanin receptor type 3 (GAL3-R) (GALR3).
GN   GALR3 OR GALNR3.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98389766; PubMed=9722565;
RA   Smith K.E., Walker M.W., Artymyshyn R., Bard J., Borowsky B.,
RA   Tamm J.A., Yao W.-J., Vaysse P.J.-J., Branchek T.A., Gerald C.,
RA   Jones K.A.;
RT   "Cloned human and rat galanin GALR3 receptors: pharmacology and
RT   activation of G-protein inwardly rectifying K+ channels.";
RL   J. Biol. Chem. 273:23321-23326(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Uterus;
RA   Bennett M.M., Lescoe M.K., Gallipoli P.Z., Ramabhadran T.V.;
RT   "Homologue of the human galanin 2 receptor gene isolated from a human
RT   uterus cDNA library.";
RL   Submitted (MAY-1998) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99048961; PubMed=9832121;
RA   Kolakowski L.F. Jr., O'Neill G.P., Howard A.D., Broussard S.R.,
RA   Sullivan K.A., Feighner S.D., Sawzdargo M., Nguyen T., Kargman S.,
RA   Shiao L.-L., Hreniuk D.L., Tan C.P., Evans J., Abramovitz M.,
RA   Chateauneuf A., Coulombe N., Ng G., Johnson M.P., Tharian A.,
RA   Khoshbouei H., George S.R., Smith R.G., O'Dowd B.F.;
RT   "Molecular characterization and expression of cloned human galanin
RT   receptors GALR2 and GALR3.";
RL   J. Neurochem. 71:2239-2251(1998).
RN   [4]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99127120; PubMed=9928159;
RA   Iismaa T.P., Fathi Z., Hort Y.J., Iben L.G., Dutton J.L., Baker E.,
RA   Sutherland G.R., Shine J.;
RT   "Structural organization and chromosomal localization of three human
RT   galanin receptor genes.";
RL   Ann. N.Y. Acad. Sci. 863:56-63(1998).
RN   [5]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20057165; PubMed=10591208;
RA   Dunham I., Hunt A.R., Collins J.E., Bruskiewich R., Beare D.M.,
RA   Clamp M., Smink L.J., Ainscough R., Almeida J.P., Babbage A.,
RA   Bagguley C., Bailey J., Barlow K., Bates K.N., Beasley O., Bird C.P.,
RA   Blakey S., Bridgeman A.M., Buck D., Burgess J., Burrill W.D.,
RA   Burton J., Carder C., Carter N.P., Chen Y., Clark G., Clegg S.M.,
RA   Cobley V., Cole C.G., Collier R.E., Connor R.E., Conroy D., Corby N.,
RA   Coville G.J., Cox A.V., Davis J., Dawson E., Dhami P.D., Dockree C.,
RA   Dodsworth S.J., Durbin R.M., Ellington A., Evans K.L., Fey J.M.,
RA   Fleming K., French L., Garner A.A., Gilbert J.G.R., Goward M.E.,
RA   Grafham D., Griffiths M.N., Hall C., Hall R., Hall-Tamlyn G.,
RA   Heathcott R.W., Ho S., Holmes S., Hunt S.E., Jones M.C., Kershaw J.,
RA   Kimberley A., King A., Laird G.K., Langford C.F., Leversha M.A.,
RA   Lloyd C., Lloyd D.M., Martyn I.D., Mashreghi-Mohammadi M.,
RA   Matthews L., Mccann O.T., Mcclay J., Mclaren S., Mcmurray A.A.,
RA   Milne S.A., Mortimore B.J., Odell C.N., Pavitt R., Pearce A.V.,
RA   Pearson D., Phillimore B.J., Phillips S.H., Plumb R.W., Ramsay H.,
RA   Ramsey Y., Rogers L., Ross M.T., Scott C.E., Sehra H.K., Skuce C.D.,
RA   Smalley S., Smith M.L., Soderlund C., Spragon L., Steward C.A.,
RA   Sulston J.E., Swann R.M., Vaudin M., Wall M., Wallis J.M.,
RA   Whiteley M.N., Willey D., Williams L., Williams S., Williamson H.,
RA   Wilmer T.E., Wilming L., Wright C.L., Hubbard T., Bentley D.R.,
RA   Beck S., Rogers J., Shimizu N., Minoshima S., Kawasaki K., Sasaki T.,
RA   Asakawa S., Kudoh J., Shintani A., Shibuya K., Yoshizaki Y., Aoki N.,
RA   Mitsuyama S., Roe B.A., Chen F., Chu L., Crabtree J., Deschamps S.,
RA   Do A., Do T., Dorman A., Fang F., Fu Y., Hu P., Hua A., Kenton S.,
RA   Lai H., Lao H.I., Lewis J., Lewis S., Lin S.-P., Loh P., Malaj E.,
RA   Nguyen T., Pan H., Phan S., Qi S., Qian Y., Ray L., Ren Q., Shaull S.,
RA   Sloan D., Song L., Wang Q., Wang Y., Wang Z., White J., Willingham D.,
RA   Wu H., Yao Z., Zhan M., Zhang G., Chissoe S., Murray J., Miller N.,
RA   Minx P., Fulton R., Johnson D., Bemis G., Bentley D., Bradshaw H.,
RA   Bourne S., Cordes M., Du Z., Fulton L., Goela D., Graves T.,
RA   Hawkins J., Hinds K., Kemp K., Latreille P., Layman D., Ozersky P.,
RA   Rohlfing T., Scheet P., Walker C., Wamsley A., Wohldmann P., Pepin K.,
RA   Nelson J., Korf I., Bedell J.A., Hillier L., Mardis E., Waterston R.,
RA   Wilson R., Emanuel B.S., Shaikh T., Kurahashi H., Saitta S.,
RA   Budarf M.L., McDermid H.E., Johnson A., Wong A.C.C., Morrow B.E.,
RA   Edelmann L., Kim U.J., Shizuya H., Simon M.I., Dumanski J.P.,
RA   Peyrard M., Kedra D., Seroussi E., Fransson I., Tapia I., Bruder C.E.,
RA   O'Brien K.P., Wilkinson P., Bodenteich A., Hartman K., Hu X.,
RA   Khan A.S., Lane L., Tilahun Y., Wright H.;
RT   "The DNA sequence of human chromosome 22.";
RL   Nature 402:489-495(1999).
CC   -!- FUNCTION: RECEPTOR FOR THE HORMONE GALANIN.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
```

```
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; AF073799; AAC35944.1; -.
DR   EMBL; AF067733; AAC18860.1; -.
DR   EMBL; AF129514; AAD47348.2; -.
DR   EMBL; AF129513; AAD47348.2; JOINED.
DR   EMBL; Z97630; CAB42831.1; -.
DR   MIM; 603692; -.
DR   InterPro; IPR000276; GPCR_Rhodpsn.
DR   Pfam; PF00001; 7tm_1; 1.
DR   PRINTS; PR00237; GPCRRHODOPSN.
DR   PROSITE; PS00237; G_PROTEIN_RECEP_F1_1; 1.
DR   PROSITE; PS50262; G_PROTEIN_RECEP_F1_2; 1.
KW   G-protein coupled receptor; Transmembrane; Glycoprotein;
KW   Multigene family; Lipoprotein; Palmitate.
FT   DOMAIN        1     20       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     21     41       1 (POTENTIAL).
FT   DOMAIN       42     57       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     58     78       2 (POTENTIAL).
FT   DOMAIN       79     96       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     97    118       3 (POTENTIAL).
FT   DOMAIN      119    138       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    139    159       4 (POTENTIAL).
FT   DOMAIN      160    184       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    185    205       5 (POTENTIAL).
FT   DOMAIN      206    236       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    237    257       6 (POTENTIAL).
FT   DOMAIN      258    259       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    260    280       7 (POTENTIAL).
FT   DOMAIN      281    368       CYTOPLASMIC (POTENTIAL).
FT   CARBOHYD      6      6       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   DISULFID     95    172       BY SIMILARITY.
FT   LIPID       308    308       PALMITATE (BY SIMILARITY).
FT   DOMAIN      218    225       POLY-ALA.
FT   DOMAIN      310    318       POLY-ARG.
SQ   SEQUENCE   368 AA;  39573 MW;  E8F46B66422C3150 CRC64;


  Query Match             3.8%;  Score 110;  DB 1;  Length 368;
  Best Local Similarity   21.9%;  Pred. No. 0.71;
  Matches   94;  Conservative   35;  Mismatches  145;  Indels  156;  Gaps   17;

Qy      87 PFIWALSLGILLLSLFLIPRAG---WLAGLLCPDPRPLE-----LALLILGVGLLDFCGQV 138
           | ::||       :||:   |    || || | |   :       | || : : | |  :
Db      21 PVVFAL-------IFLLGTVGNGLVLAVLLQPGPSAWQEPGSTTDLFILNLAVADLCFIL 73

Qy     139 CFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPAIDWDTSALAPYLGTQEEC 198
           |  | :| :                              | | |  |
Db      74 CCVPFQATI----------------------------YTLDA--W--------------- 88

Qy     199 LFGLLT------LIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPHCCPCRARLAFRN 252
           |||  |       ||:||  |::                 |:| |:  :       || |
Db      89 LFGALVCKAVHLLIYLTMYASSFT--------------LAAVSVDRY-------LAVR- 125

Qy     253 LGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGLYQGVPRAEPGTE 312
                 | |  |  || |    |    |:    |:  |  : |   |  :    |  |
Db     126 --------HPLRSRALRTPRNARAAVGLVWLLAALFSAPYLSYYGTVRYGALELCVPAWE 177

Qy     313 ARRHYDEGVRMGSLGLFLQCA-ISLVFSLVMDRLVQRFG----------TRAVYLASVAA 361
           |      |   : |  |  | :|| :   :  |    |            ||   |  |
Db     178 DARRRALDVATFAAGYLLPVAVVSLAYGRTLRFLWAAVGPAGAAAAAEARRRATGRAGRAM 237

Qy     362 FPVAA------------------GATCLSHSVAVVTASAALTGFTFSALQILPYTLASLY 403
           |||                   |    | :      ::    :  | |  | | ||| :
Db     238 LAVAALYALCWGPHHALILCFWYGRFAFSPATYACRLASHCLAYANSCLNPLVYALASRH 297

Qy     404 ----------------HREKQVFLPKYRGDTGGASSEDSLMTSFLPGPKPGAPFPNGHVG 447
                           || ::  | : |  : |            |   ||   |:| :
Db     298 FRARFRRLWPCGRRRRHRARRA-LRRVRPASSG-----------PPGCPGDARPSGRLL 344
```

```
Qy     448 AGGSGLLPPP 457
           ||| |  | |
Db     345 AGG-GQGPEP 353


RESULT  15
NQOC_THETH
ID   NQOC_THETH     STANDARD;      PRT;   606 AA.
AC   Q56227;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   NADH-ubiquinone oxidoreductase chain 12 (EC 1.6.5.3) (NADH
DE   dehydrogenase 1, chain 12) (NDH-1, chain 12).
GN   NQO12.
OS   Thermus aquaticus (subsp. thermophilus).
OC   Bacteria; Thermus/Deinococcus group; Thermus group; Thermus.
OX   NCBI_TaxID=274;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=HB8 / ATCC 27634;
RX   MEDLINE=97172490; PubMed=9020134;
RA   Yano T., Chu S.S., Sled V.D., Ohnishi T., Yagi T.;
RT   "The proton-translocating NADH-quinone oxidoreductase (NDH-1) of
RT   thermophilic bacterium Thermus thermophilus HB-8. Complete DNA
RT   sequence of the gene cluster and thermostable properties of the
RT   expressed NQO2 subunit.";
RL   J. Biol. Chem. 272:4201-4211(1997).
CC   -!- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.
CC   -!- SUBUNIT: COMPOSED OF 14 DIFFERENT SUBUNITS. SUBUNITS NQO7-14
CC       CONSTITUTE THE MEMBRANE SECTOR OF THE COMPLEX.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- SIMILARITY: TO POLYPEPTIDE 5 OF THE NADH-UBIQUINOL OXIDOREDUCTASE
CC       OF CHLOROPLASTS OR MITOCHONDRIA.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; U52917; AAA97949.1; -.
DR   InterPro; IPR003916; NADHub_oxdrdctse5.
DR   InterPro; IPR001750; Oxidored_q1.
DR   InterPro; IPR001516; Oxidored_q1_N.
DR   Pfam; PF00361; oxidored_q1; 1.
DR   Pfam; PF00662; oxidored_q1_N; 1.
DR   PRINTS; PR01434; NADHDHGNASE5.
KW   Oxidoreductase; NAD; Ubiquinone; Transmembrane.
FT   TRANSMEM      1     21       POTENTIAL.
FT   TRANSMEM     29     49       POTENTIAL.
FT   TRANSMEM     75     95       POTENTIAL.
FT   TRANSMEM    104    124       POTENTIAL.
FT   TRANSMEM    126    146       POTENTIAL.
FT   TRANSMEM    165    185       POTENTIAL.
FT   TRANSMEM    206    226       POTENTIAL.
FT   TRANSMEM    234    254       POTENTIAL.
FT   TRANSMEM    266    286       POTENTIAL.
FT   TRANSMEM    303    323       POTENTIAL.
FT   TRANSMEM    325    345       POTENTIAL.
FT   TRANSMEM    364    384       POTENTIAL.
FT   TRANSMEM    397    417       POTENTIAL.
FT   TRANSMEM    441    461       POTENTIAL.
FT   TRANSMEM    492    512       POTENTIAL.
FT   TRANSMEM    584    604       POTENTIAL.
SQ   SEQUENCE   606 AA;  65141 MW;  47318E8F2665295C CRC64;


  Query Match             3.8%;  Score 110;  DB 1;  Length 606;
  Best Local Similarity   21.4%;  Pred. No. 1.2;
```

Matches 126; Conservative 69; Mismatches 198; Indels 196; Gaps 27;

```
Qy      62 VLGLVCVPLLGSASDHWRGRYGR--RRPFIWALSLGILLSLFLIP-----------RAGW 108
           :|| : :|||| |     | :|:  | |    |: |::|: ||:           :| |
Db       3 LLGTILLPLLGFA---LLGLFGKRMREPLPGVLASGLVLASFLLGAGLLLSGGARFQAEW 59

Qy     109 LAGL---LCPDPRPLELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVY- 164
           | |:   |  |     : |:: ||| |     : :         :  || : |  :  |
Db      60 LPGIPFSLLLDNLSGFMLLIVTGVGFLIHVYAIGY---------MGGDPGYSR--FFAYF 108

Qy     165 ----AFMISLGGCLGYLLPAIDWDTSALAPYLGTQEECLFG------------------ 201
               | |::|     | :  | |:   || :|      | |
Db     109 NLFIAMMLTLVLADSYPVMFIGWEGVGLASFL------LIGFWYKNPQYADSARKAFIVN 162

Qy     202 -------LLTLIFLTCVAATLLVAE--EAALGPTEPAEGLSAPSLSPHCCPCRARLAFRN 252
                  :| :  |  :  || ::|  ||  ||      |  | |          ||
Db     163 RIGDLGFMLGMAILWALYGTLSISELKEAMEGP------LKNPDL----------LALAG 206

Qy     253 LGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEG----LYQGVPRAE 308
           |                                    | :   ||:     |   :| |
Db     207 L-----------------------------------LLFLGAVGKSAQIPLMVWLPDAM 230

Qy     309 PGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVAAFPVAA-G 367
           |          | : |::|    | ::|:: |     :    | | : |   ::| |
Db     231 AGPTPVSALIHAATMVTAGVYLIARSSFLYSVLPD---VSYAIAVVGLLTAAYGALSAFG 287

Qy     368 ATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVF----------------- 410
           | :   ||  | |    |: | |  :  | :| |:|     |
Db     288 QTDIKKIVAYSTISQ--LGYMFLAAGVGAYWVA-LFHVFTHAFFKALLFLASGSVIHALG 344

Qy     411 -------------LPKYRGDT-------------GGASSEDSLMTSFLPGPKPGAPFPN 443
                       ||: |                 |  |:|::: : |  |  |  |
Db     345 GEQDVRKMGGLWKHLPQTRWHALIGALALGGLPLLSGFWSKDAILAATLTYPFGGVGF-- 402

Qy     444 GHVGAGGSGLLPPPPALCGASACDVSVRVVVGE------PTEARVV---PGRGICLDLAI 494
           :|||    ||     |    |  | :||       | ||  |   |   : |  ::
Db     403 -YVGA----LL--VAVLTAMYAMRWFVLVFLGEERGHHHPHEAPPVMLWPNHLLALG-SV 454

Qy     495 LDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVS-AAGLGLVAIYFATQVV 542
           |     |    |::    :       | |: :|  |  ||:|:  |  ::
Db     455 LAGYLALPHPLPNVLEPFLKPALAEVEAHHLSLGAEWGLIALSAAVALL 503
```

Search completed: June 26, 2002, 13:48:18
Job time: 401 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5

OM protein - protein search, using sw model

Run on: June 26, 2002, 12:40:07 ; Search time 70.32 Seconds
(without alignments)
755.651 Million cell updates/sec

Title: US-09-838-785-2
Perfect score: 2861
Sequence: 1 MVQRLWVSRLLRHRKAQLLL.....AIYFATQVVFDKSDLAKYSA 553

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_71:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	347.5	12.1	515	2	T14340	sucrose-proton tra
2	337.5	11.8	523	2	T12198	sucrose transport
3	330	11.5	516	2	JQ2389	sucrose transport
4	328.5	11.5	594	2	G84441	probable sucrose-p
5	324.5	11.3	525	2	S28052	sucrose transport
6	323.5	11.3	512	2	S38196	sucrose transport
7	323.5	11.3	512	2	G86360	probable sucrose-p
8	317.5	11.1	507	2	S48789	sucrose transport
9	311	10.9	533	2	S43142	sucrose transport
10	305.5	10.7	501	2	T14339	sucrose-proton tra
11	297.5	10.4	428	2	S48788	sucrose transport
12	294	10.3	510	2	S38657	sucrose transport
13	294	10.3	523	2	S51114	sucrose-proton sym
14	292.5	10.2	512	2	F96741	probable sucrose t
15	292	10.2	537	2	T02982	probable sucrose t
16	286	10.0	474	2	A86234	hypothetical prote
17	281.5	9.8	491	2	E96691	probable sucrose-p
18	279.5	9.8	513	2	S38197	sucrose transport
19	276.5	9.7	492	2	A84520	probable sucrose-p
20	273	9.5	503	2	S52377	sucrose transport
21	235.5	8.2	553	2	T38541	probable sucrose c
22	191	6.7	452	2	F75217	hypothetical prote
23	187.5	6.6	541	2	B87532	transporter, proba
24	142	5.0	544	2	S75696	melibiose carrier
25	139.5	4.9	454	2	A75444	hypothetical prote
26	138.5	4.8	430	2	E75217	transporter PAB217
27	136	4.8	418	2	B87536	membrane protein,
28	133	4.6	389	2	G83413	probable MFS trans
29	128.5	4.5	422	2	G83503	probable MFS trans
30	128.5	4.5	594	2	A83096	probable permease
31	127.5	4.5	472	2	F82639	resistance protein
32	127	4.4	451	2	E81781	probable integral
33	126	4.4	394	2	G98213	drug efflux protei
34	126	4.4	394	2	AB3073	tetracycline resis
35	126	4.4	451	2	A81206	sugar transporter,
36	125.5	4.4	399	2	JQ1479	tetracycline resis
37	125	4.4	391	2	S74688	hypothetical prote
38	125	4.4	407	2	H75515	tetracycline-efflu
39	124.5	4.4	399	1	S38656	tetA protein - Pse
40	123.5	4.3	400	2	D97643	probable efflux pr
41	123.5	4.3	400	2	AG2866	MFS permease [drug
42	123	4.3	440	2	AD0986	hypothetical metab
43	121.5	4.2	398	2	C90349	multidrug-efflux t
44	121.5	4.2	503	2	S43017	puromycin resistan
45	121	4.2	483	2	AF2363	hypothetical prote

ALIGNMENTS

RESULT 1
T14340
sucrose-proton transport protein - carrot
N;Alternate names: sucrose/H+ symporter protein
C;Species: Daucus carota (carrot)
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 21-Jul-2000
C;Accession: T14340
R;Shakya, R.; Sturm, A.
Plant Physiol. 118, 1473-1480, 1998
A;Title: Characterization of source- and sink-specific sucrose/H+ symporters from car
A;Reference number: Z17991; MUID:99063785
A;Accession: T14340
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-515 <SHA>
A;Cross-references: EMBL:Y16768; NID:g2969883; PIDN:CAA76369.1; PID:g2969884
A;Experimental source: cultivar Namtaise; root
C;Genetics:
A;Note: SUT2
C;Superfamily: common tobacco sucrose transport protein

Query Match 12.1%; Score 347.5; DB 2; Length 515;
Best Local Similarity 25.9%; Pred. No. 1e-18;
Matches 131; Conservative 80; Mismatches 215; Indels 79; Gaps 17;

```
Qy      17 QLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIGPVLGLVCVPLLGSASD 76
           :|:||  :  |::  |  :: : | :  :|:  |:   :   ||: |::  |::|  ||
Db      34 KLVLVAAIAAGVQFGWALQLSLLTPYVQLLGIPHKWAAYIWLCGPISGMLVQPIVGYYSD 93

Qy      77 HWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGL-------LCPDPRPLELALLILGV 129
           | :  :|||||| : :  : :|: ||   |: | :       :    :|  : : ::|
Db      94 HCQSSFGRRRPFIASGAGCVAISVILI---GFAADISYKAGDDMSKTLKPRAVTVFVIGF 150

Qy     130 GLLDFCGQVCFTPLEALLSDLFR-DPDHCRQAYSVYAFMISLGGCLGY----------LL 178
           :||   :  |  |||:||  |   | | : |:| :::|  |||          |
Db     151 WILDVANNMLQGPCRALLADLCSGDTRRMRSANAFYSFFMAVGNILGYAAGSYNNLYKLF 210

Qy     179 PAIDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLS 238
           |    | |   |    : |    : |: :  | |  :| | :  || : |:    |
Db     211 PFS--KTHACDLYCANLKSCFIISIALLIIITVVALSVVRENS--GPPDDADAAEEP--- 263

Qy     239 PHCCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGE 298
           |   ::    | ||  |  |   || :  | :   :|:|   | || ||::|
Db     264 ----PSSGKIPV--FGELLGALKDL----PRPMLLLLIVTCLNWIAWFPFILFDTDWMGR 313

Qy     299 GLYQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLAS 358
           :| |   |  |    : ||:||| |:||| |   :  : |: :: ||:  |   : |
Db     314 EIYGGT--AGQG----KLYDQGVRAGALGLLLNSVVLGLTSIAVEYLVRGVGGVKI-LWG 366

Qy     359 VAAFPVAAGATCL----------SHSV--------AVVTASA----ALTGFTFSALQIL 395
           | :| |               ||        | | | |    :: |   |    :
```

```
Db    367 FVNFILAIGLVMTVVVSKVAQHQREHSANGQLLPPSAGVKAGALSLFSILGIPLSITYSI 426

Qy    396 PYTLASLYHREKQVFLPKYRGDTGGASSEDSLMTSFLPGPKPG-------APFPNGHVGA 448
          |: |||:|            |    |     :: | | ||           |  | : |
Db    427 PFALASIYSSGSGAGQGLSLGVLNLAIVVPQMIVSVLAGPFDSLFGGGNLPAFVVGAISA 486

Qy    449 GGSGLLP----PPPALCGASACDVS 469
            ||:|     | |:   ||   :|
Db    487 AISGVLAIVLLPKPSKDAASKLSLS 511


RESULT   2
T12198
sucrose transport protein - fava bean
C;Species: Vicia faba (fava bean)
C;Date: 23-Jul-1999 #sequence_revision 23-Jul-1999 #text_change 20-Jun-2000
C;Accession: T12198
R;Weber, H.; Borisjuk, L.; Heim, U.; Sauer, N.; Wobus, U.
Plant Cell 9, 895-908, 1997
A;Title: A role for sugar transporters during seed development: molecular characterizati
A;Reference number: Z17451; MUID:97355984
A;Accession: T12198
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-523 <WEB>
A;Cross-references: EMBL:Z93774
C;Genetics:
A;Gene: sut
C;Superfamily: common tobacco sucrose transport protein

  Query Match             11.8%;  Score 337.5;  DB 2;  Length 523;
  Best Local Similarity   25.8%;  Pred. No. 6e-18;
  Matches  134;  Conservative   76;  Mismatches  185;  Indels  125;  Gaps   20;

Qy     17 QLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIGPVLGLVCVPLLGSASD 76
          ::::|  :  |::   |  :: : | :  :|:   :   :   ||: |::  |::|  ||
Db     36 KIMVVASIAAGVQFGWALQLSLLTPYVQLLGIHHTWAAYIWLCGPISGMLVQPIVGYHSD 95

Qy     77 HWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGL-------LCPDPRPLELALLILGV 129
              |:|||||| | |: : :::|||   |: | |       |    ||  : : ::|
Db     96 RCTSRFGRRRPFIAAGSIAVAIAVFLI---GYAADLGHSFGDSLDQKVRPRAIGIFVVGF 152

Qy    130 GLLDFCGQVCFTPLEALLSDLFR-DPDHCRQAYSVYAFMISLGGCLGYLLPAID------ 182
           :||    :   |  |||| ||   :    | | : ::| :::|  |||   |
Db    153 WILDVANNMLQGPCRALLGDLCAGNQRKTRNANAFFSFFMAVGNVLGYAAGAYSKLYHVF 212

Qy    183 --WDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLV-AEEAALGPTEPA----EGLSAP 235
            | |   |    : | |  |::  || :| : |:  :| || | :       :| |:
Db    213 PFTKTKACNVYCANLKSCFF--LSIALLTVLATSALIYVKETALTPEKTVVTTEDGGSSG 270

Qy    236 SLSPHCCPCRARL--AFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYT 293
          :    ||  :|  ||: |                | :  | :    :|:|   | || |
Db    271 GM-----PCFGQLSGAFKEL-------------KRPMWILLLVTCLNWIAWFPFLLFDT 311

Qy    294 DFVGEGLYQGVPRAEPGTEARRH-YDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTR 352
          |::|: :|        ||    | || ||| |:||| |   :    || :| | :  |
Db    312 DWMGKEVY-------GGTVGEGHAYDMGVREGALGLMLNSVVLGATSLGVDILARGVG-G 363

Qy    353 AVYLASVAAFPVAAGATCLSHSVAVV-------------------------TASAALTGF 387
            |  :  | :   | ||  :| |                           : ||| |
Db    364 VKRLWGIVNFLL---AICLGLTVLVTKLAQHSRQYAPGTGALGDPLPPSEGIKAGALTLF 420

Qy    388 TFSALQI-----LPYTLASLYHREKQVFLPKYRGDTGGASSEDSL------------MTS 430
          :  : :     :|: |||::             | ||    ||             |
Db    421 SVLGVPLAITYSIPFALASIF------------SSTSGAGQGLSLGVLNLAIVIPQMFVS 468

Qy    431 FLPGPKPG-------APFPNGHVGAGGSG-----LLPPPP 458
          | ||            |  | | |  ||     ||| ||
Db    469 VLSGPWDALFGGGNLPAFVVGAVAALASGILSIILLPSPP 508


RESULT   3
```

```
JQ2389
sucrose transport protein - potato
C;Species: Solanum tuberosum (potato)
C;Date: 28-Aug-1985 #sequence_revision 07-Oct-1994 #text_change 21-Jul-2000
C;Accession: JQ2389; S40310
R;Riesmeier, J.W.; Hirner, B.; Frommer, W.B.
Plant Cell 5, 1591-1598, 1993
A;Title: Potato sucrose transporter expression in minor veins indicates a role in phl
A;Reference number: JQ2389; MUID:94146554
A;Accession: JQ2389
A;Molecule type: mRNA
A;Residues: 1-516 <RIE>
A;Cross-references: EMBL:X69165; NID:g439293; PIDN:CAA48915.1; PID:g439294
A;Experimental source: cv. Desiree
C;Comment: The gene encoding for this protein is highly expressed in mature leaves.
C;Superfamily: common tobacco sucrose transport protein
C;Keywords: glycoprotein; transmembrane protein
F;31-53/Domain: transmembrane #status predicted <TM1>
F;67-86/Domain: transmembrane #status predicted <TM2>
F;103-122/Domain: transmembrane #status predicted <TM3>
F;141-160/Domain: transmembrane #status predicted <TM4>
F;180-200/Domain: transmembrane #status predicted <TM5>
F;226-248/Domain: transmembrane #status predicted <TM6>
F;285-304/Domain: transmembrane #status predicted <TM7>
F;331-349/Domain: transmembrane #status predicted <TM8>
F;366-385/Domain: transmembrane #status predicted <TM9>
F;409-427/Domain: transmembrane #status predicted <TM10>
F;429-448/Domain: transmembrane #status predicted <TM11>
F;3,92/Binding site: carbohydrate (Asn) (covalent) #status predicted

  Query Match             11.5%;  Score 330;  DB 2;  Length 516;
  Best Local Similarity   24.7%;  Pred. No. 2.2e-17;
  Matches  132;  Conservative   93;  Mismatches  197;  Indels  112;  Gaps   21;

Qy      4 RLWVSRLLRHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIGPVL 63
          :||          ::::|  :  |::   |  :: : | :  :|:  || : :   ||:
Db     29 KLW----------KIIVVASIAAGVQFGWALQLSLLTPYVQLLGIPHKFASFIWLCGPIS 78

Qy     64 GLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGL-------LCPD 116
          |::  |::|  ||:   |:|||||| | :  :::::|||   |: | |       |
Db     79 GMIVQPVVGYYSDNCSSRFGRRRPFIAAGAALVMIAVFLI---GFAADLGHASGDTLGKG 135

Qy    117 PRPLELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPD-HCRQAYSVYAFMISLGGCLG 175
           :|  :|: ::|  :||    :   |  |||:||        | | : ::| :::|  ||
Db    136 FKPRAIAVFVVGFWILDVANNMLQGPCRALLADLSGGKSGRMRTANAFFSFFMAVGNILG 195

Qy    176 YLLPAIDW--------DTSALAPYLGTQEECLF-GLLTLIFLTCVAATLLVAEE-AALGP 225
          |   :              | |   |    : | |  :  |: || :| ||:   |
Db    196 YAAGSYSHLFKVFPFSKTKACDMYCANLKSCFFIAIFLLLSLTTIALTLVRENELPEKDE 255

Qy    226 TEPAEGLSAPSLSPHCCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMAL 285
          |  | |:           :::: |   | :   | :|    || :  | :    :|:|
Db    256 QEIDEKLAGAG--------KSKVPF--FGEIFGALKEL----PRPMWILLLVTCLNWIAW 301

Qy    286 MTFTLFYTDFVGEGLYQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRL 345
          | |: ||:: : :: |    : |    | || ||| |::|| ||  :    || :: |
Db    302 FPFFLYDTDWMAKEVFGG----QVGD--ARLYDLGVRAGAMGLLLQSVVLGFMSLGVEFL 355

Qy    346 VQRFGTRAVYLASVAAFPVAAGATCLSHSVAVV---------TASAALTGFT-------- 388
          :: |  |  |  :  | :   | ||: :: |           .   :  | | |
Db    356 GKKIG-GAKRLWGILNFVL---AICLAMTILVTKMAEKSRQHDPAGTLMGPTPGVKIGAL 411

Qy    389 --FSALQI-------LPYTLASLYHREK------------QVFLPKYRGDTGGASSEDS 426
            |:|| |        :|: |||::   :              : :|:       |   :|
Db    412 LLFAALGIPLAATFSIPFALASIFSSNRGSGQGLSLGVLNLAIVVPQMLVSLVGGPWDDL 471

Qy    427 LMTSFLPGPKPGAPFPNGHVGAGGSG-----LLPPPPALCGASACDVSVRVVVG 475
             |||      |  | | |  |       :|| |||       |    | :|
Db    472 FGGGNLPG------FVVGAVAAAASAVLALTMLPSPPA-------DAKPAVAMG 512


RESULT   4
```

```
G84441
probable sucrose-proton symporter [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2001
C;Accession: G84441
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487
A;Accession: G84441
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-594 <STO>
A;Cross-references: GB:AE002093; NID:g3461813; PIDN:AAC32907.1; GSPDB:GN00139
C;Genetics:
A;Gene: At2g02860
A;Map position: 2
C;Superfamily: common tobacco sucrose transport protein

  Query Match             11.5%;  Score 328.5;  DB 2;  Length 594;
  Best Local Similarity   24.1%;  Pred. No. 3.3e-17;
  Matches  120;  Conservative   72;  Mismatches  174;  Indels  131;  Gaps   18;

Qy      22 NLLTFGLEVCLAAGITY--------VPPLLLEVGVEEKFMTMVLGIGPVLGLVCVPLLGS 73
           :|:|  |   :|||: :        : | :  :|:   | : :   ||: |||  | :|
Db      59 SLVTLVLSCTVAAGVQFGWALQLSLLTPYIQTLGISHAFSSFIWLCGPITGLVVQPFVGI 118

Qy      74 ASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLA--GLLCPD----------PRPLE 121
             ||    :|||||||   |  | ::: :|   |: |  | |  |           |
Db     119 WSDKCTSKYGRRRPFILVGSFMISIAVIII---GFSADIGYLLGDSKEHCSTFKGTRTRA 175

Qy     122 LALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPAI 181
             : |:|  |||        |  |||:||   ||    | :|:   :::|  ||:   |
Db     176 AVVFIIGFWLLDLANNTVQGPARALLADL-SGPDQRNTANAVFCLWMAIGNILGFSAGAS 234

Qy     182 ----DWDTSALAPYLGTQEECLFG--------LLTLIFLT-CVAATLLVAEEAALGPTEP 228
               :|      |:| |   |              || ::||| |   |:  |:|      :|
Db     235 GKWQEW-----FPFL-TSRACCAACGNLKAAFLLAVVFLTICTLVTIYFAKEIPFTSNKP 288

Qy     229 AE-GLSAPSLSPHCCPCRARLAFRNL---------------------------------- 253
              |||  |          |  : |
Db     289 TRIQDSAPLLDD--------LQSKGLEHSKLNNGTANGIKYERVERDTDEQFGNSENEHQ 340

Qy     254 --------GALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGLYQGVP 305
                  |::|  |       :|  :  : :    :|::   | || ||::|  :| | |
Db     341 DETYVDGPGSVLVNLLTSLRHLPPAMHSVLIVMALTWLSWFPFFLFDTDWMGREVYHGDP 400

Qy     306 RAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVY-LASVAAFPV 364
             :  :     ||:||| |:||| |   :  : | ::: : || | | |: |::   |
Db     401 TGD--SLHMELYDQGVREGALGLLLNSVVLGISSFLIEPMCQRMGARVVWALSNFTVFAC 458

Qy     365 AAGATCLS-------------------HSVAVVTASAALTGFTFSALQILPYTLASLY 403
           ||   :|                    : ||:    || ||  :    :|::: :
Db     459 MAGTAVISLMSLSDDKNGIEYIMRGNETTRTAAVIV--FALLGFPLAITYSVPFSVTA-- 514

Qy     404 HREKQVFLPKYRGDTGG 420
                 :   |:||
Db     515 ---------EVTADSGG 522


RESULT   5
S28052
sucrose transport protein - spinach
N;Alternate names: sucrose carrier protein; sucrose permease
C;Species: Spinacia oleracea (spinach)
C;Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 21-Jul-2000
C;Accession: S28052
R;Riesmeier, J.W.; Willmitzer, L.; Frommer, W.B.
EMBO J. 11, 4705-4713, 1992
A;Title: Isolation and characterization of a sucrose carrier cDNA from spinach by functi
A;Reference number: S28052; MUID:93099843
A;Accession: S28052
A;Molecule type: mRNA
A;Residues: 1-525 <RIE>
A;Cross-references: EMBL:X67125; NID:g21318; PIDN:CAA47604.1; PID:g21319
C;Superfamily: common tobacco sucrose transport protein
C;Keywords: transmembrane protein

  Query Match             11.3%;  Score 324.5;  DB 2;  Length 525;
  Best Local Similarity   23.2%;  Pred. No. 5.8e-17;
  Matches  129;  Conservative   91;  Mismatches  196;  Indels  141;  Gaps   20;

Qy      23 LLTFGLEVCLAAGITY--------VPPLLLEVGVEEKFMTMVLGIGPVLGLVCVPLLGSA 74
           |   ||   :|||: :        : | :  :|:   :   :   ||: |::  ||:|
Db      35 LKKLGLVASVAAGVQFGWALQLSLLTPYVQLLGIPHTWAAYIWLCGPISGMIVQPLVGYY 94

Qy      75 SDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLA--GLLCPDP-----RPLELALLIL 127
           ||    |:|||||||  |  :  : ::: ||   |: |  |    ||     :|  :|: ::
Db      95 SDRCTSRFGRRRPFIAAGAALVAVAVGLI---GFAADIGAASGDPTGNVAKPRAIAVFVV 151

Qy     128 GVGLLDFCGQVCFTPLEALLSDLFR-DPDHCRQAYSVYAFMISLGGCLGYLLPAID---- 182
           |  :||        |  |||:|:         | | : ::| ::||   ||   :
Db     152 GFWILDVANNTLQGPCRALLADMAAGSQTKTRYANAFFSFFMALGNIGGYAAGSYSRLYT 211

Qy     183 ----WDTSALAPYLGTQEECLFGLLT-LIFLTCVAATLLVAEEAALGPTEPAEGLSAPSL 237
                 |:|    |    : | |  :| || || :| :::   :  :   :  | |   :
Db     212 VFPFTKTAACDVYCANLKSCFFISITLLIVLTILALSVVKERQITIDEIQEEEDLKNRNN 271

Qy     238 SPHCCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVG 297
           |  |    ||| |   | |:  |  |    |: :  | :    :|:|   | || ||::|
Db     272 SSGC----ARLPF--FGQLIGALKDL----PKPMLILLLVTALNWIAWFPFLLFDTDWMG 321

Qy     298 EGLYQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFG--TRAVY 355
           : :| |           : ||:||  |:||| :   :  | || :: | :  |   |
Db     322 KEVYGGT------VGEGKLYDQGVHAGALGLMINSVVLGVMSLSIEGLARMVGGAKRLWG 375

Qy     356 LASVAAFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYR 415
           : ::        | ||: :| :|| ||                  : |:    :
Db     376 IVNIIL------AVCLAMTV-LVTKSAE------------------HFRDSHHIM---- 405

Qy     416 GDTGGASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVG 475
                                          || : |||||  |     :::  |:|
Db     406 ---------------------------------GSAVPPPPPA--GVKGGALAIFAVLG 429

Qy     476 EPTEARV-VP-----------GRGICLDLAILDSAFLLSQV----------------AP 506
           |      :|           | |  | | :|: | :: |:                  |
Db     430 IPLAITFSIPFALASIFSASSGSGQGLSLGVLNLAIVVPQMFVSVTSGPWDAMFGGGNLP 489

Qy     507 SLFMGSIVQLSQSVTAY 523
           :  :|::   : :| ::
Db     490 AFVVGAVAATASAVLSF 506


RESULT   6
S38196
sucrose transport protein SUC2 - Arabidopsis thaliana
N;Alternate names: sucrose-proton symporter SUC2
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 01-Dec-2000
C;Accession: S38196; T00773
R;Sauer, N.K.
submitted to the EMBL Data Library, October 1993
A;Description: SUC1 and SUC2: two sucrose transporters from Arabidopsis thaliana.
A;Reference number: S38196
A;Accession: S38196
A;Molecule type: mRNA
A;Residues: 1-512 <SAU>
A;Cross-references: EMBL:X75382; NID:g407091; PID:g407092
R;Vysotskaia, V.S.; Schwartz, J.R.; Toriumi, M.; Yu, G.; Oji, O.; Kwan, A.; Liu, S.;
rtz, D.; Li, Y.; Palm, C.J.; Shinn, P.; Sun, H.; Davis, R.W.; Ecker, J.R.; Federspiel
submitted to the EMBL Data Library, July 1998
A;Description: Arabidopsis thaliana chromosome 1 BAC T22J18 sequence, complete sequen
```

A;Reference number: Z14202
A;Accession: T00773
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-490,'A',492-512 <VYS>
A;Cross-references: EMBL:AC003979; NID:g3172156; PID:g3287687; GSPDB:GN00059; ATSP:T22J1
A;Experimental source: cultivar Columbia
C;Genetics:
A;Gene: SUC2; ATSP:T22J18.12
A;Map position: 1
A;Introns: 419/3; 441/1; 455/3
C;Superfamily: common tobacco sucrose transport protein
C;Keywords: sugar transport

Query Match 11.3%; Score 323.5; DB 2; Length 512;
Best Local Similarity 24.1%; Pred. No. 6.7e-17;
Matches 123; Conservative 92; Mismatches 195; Indels 101; Gaps 18;

```
Qy      14 RKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIGPVLGLVCVPLLGS 73
           |  ::: |: :  |::   |  :: : | :  :|:  |: :::   ||: |::  |::|
Db      28 RLRKIISVSSIAAGVQFGWALQLSLLTPYVQLLGIPHKWASLIWLCGPISGMLVQPIVGY 87

Qy      74 ASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGL-------LCPDPRPLELALLI 126
            ||    |:|||||||| | :  : :::|||   |: | :       |   |:   :|:
Db      88 HSDRCTSRFGRRRPFIVAGAGLVTVAVFLI---GYAADIGHSMGDQLDKPPKTRAIAIFA 144

Qy     127 LGVGLLDFCGQVCFTPLEALLSDLFR-DPDHCRQAYSVYAFMISLGGCLGY-------LL 178
           ||  :||       |  | |:||   :    | | : ::| :::|  |||       |
Db     145 LGFWILDVANNTLQGPCRAFLADLSAGNAKKTRTANAFFSFFMAVGNVLGYAAGSYRNLY 204

Qy     179 PAIDWD-TSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSL 237
             : :  | :   |    : | |  :||: |     :|   :|    |   |:| :
Db     205 KVVPFTMTESCDLYCANLKTCFFLSITLL-LIVTFVSLCYVKEKPWTPEPTADGKA---- 259

Qy     238 SPHCCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVG 297
                 : : |   | :     :|     | :  | :    :|:|   | || ||::|
Db     260 --------SNVPF--FGEIFGAFKEL----KRPMWMLLIVTALNWIAWFPFLLFDTDWMG 305

Qy     298 EGLYQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLA 357
             :| |   |      ::: |::||| |:||| |   :    || :: : :: |  |  |
Db     306 REVYGGNSDATATAASKKLYNDGVRAGALGLMLNAIVLGFMSLGVEWIGRKLG-GAKRLW 364

Qy     358 SVAAFPVAAGATCLSHSVAV----------VTASA----ALTG----FT 388
            :  | :   | ||: :| |           |||  |    |: |     |
Db     365 GIVNFIL---AICLAMTVVVTKQAENHRRDHGGAKTGPPGNVTAGALTLFAILGIPQAIT 421

Qy     389 FSALQILPYTLASLYHREK----------QVFLPKYRGDTGGASSEDSLMTSFLPGP 435
           ||    :|: |||::                : :|:     ||   ::       :|
Db     422 FS----IPFALASIFSTNSGAGQGLSLGVLNLAIVVPQMVISVGGGPFDELFGGGNIPA- 476

Qy     436 KPGAPFPNGHVGAGGSGLL-------PPPPA 459
                 |  | : |  ||:|       ||| |
Db     477 -----FVLGAIAAAVSGVLGLTVLPSPPPDA 502
```

RESULT 7
G86360
probable sucrose-proton symporter SUC2 protein T22J18.12 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 02-Nov-2001
C;Accession: G86360
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719

A;Accession: G86360
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-512 <STO>
A;Cross-references: GB:AE005172; NID:g3287687; PIDN:AAC25515.1; GSPDB:GN00141
C;Genetics:
A;Map position: 1
C;Superfamily: common tobacco sucrose transport protein

Query Match 11.3%; Score 323.5; DB 2; Length 512;
Best Local Similarity 24.1%; Pred. No. 6.7e-17;
Matches 123; Conservative 92; Mismatches 195; Indels 101; Gaps 18;

```
Qy      14 RKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIGPVLGLVCVPLLGS 73
           |  ::: |: :  |::   |  :: : | :  :|:  |: :::   ||: |::  |::|
Db      28 RLRKIISVSSIAAGVQFGWALQLSLLTPYVQLLGIPHKWASLIWLCGPISGMLVQPIVGY 87

Qy      74 ASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGL-------LCPDPRPLELALLI 126
            ||    |:|||||||| | :  : :::|||   |: | :       |   |:   :|:
Db      88 HSDRCTSRFGRRRPFIVAGAGLVTVAVFLI---GYAADIGHSMGDQLDKPPKTRAIAIFA 144

Qy     127 LGVGLLDFCGQVCFTPLEALLSDLFR-DPDHCRQAYSVYAFMISLGGCLGY-------LL 178
           ||  :||       |  | |:||   :    | | : ::| :::|  |||       |
Db     145 LGFWILDVANNTLQGPCRAFLADLSAGNAKKTRTANAFFSFFMAVGNVLGYAAGSYRNLY 204

Qy     179 PAIDWD-TSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSL 237
             : :  | :   |    : | |  :||: |     :|   :|    |   |:| :
Db     205 KVVPFTMTESCDLYCANLKTCFFLSITLL-LIVTFVSLCYVKEKPWTPEPTADGKA---- 259

Qy     238 SPHCCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVG 297
                 : : |   | :     :|     | :  | :    :|:|   | || ||::|
Db     260 --------SNVPF--FGEIFGAFKEL----KRPMWMLLIVTALNWIAWFPFLLFDTDWMG 305

Qy     298 EGLYQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLA 357
             :| |   |      ::: |::||| |:||| |   :    || :: : :: |  |  |
Db     306 REVYGGNSDATATAASKKLYNDGVRAGALGLMLNAIVLGFMSLGVEWIGRKLG-GAKRLW 364

Qy     358 SVAAFPVAAGATCLSHSVAV----------VTASA----ALTG----FT 388
            :  | :   | ||: :| |           |||  |    |: |     |
Db     365 GIVNFIL---AICLAMTVVVTKQAENHRRDHGGAKTGPPGNVTAGALTLFAILGIPQAIT 421

Qy     389 FSALQILPYTLASLYHREK----------QVFLPKYRGDTGGASSEDSLMTSFLPGP 435
           ||    :|: |||::                : :|:     ||   ::       :|
Db     422 FS----IPFALASIFSTNSGAGQGLSLGVLNLAIVVPQMVISVGGGPFDELFGGGNIPA- 476

Qy     436 KPGAPFPNGHVGAGGSGLL-------PPPPA 459
                 |  | : |  ||:|       ||| |
Db     477 -----FVLGAIAAAVSGVLGLTVLPSPPPDA 502
```

RESULT 8
S48789
sucrose transport protein - common tobacco
C;Species: Nicotiana tabacum (common tobacco)
C;Date: 27-Jan-1995 #sequence_revision 27-Jan-1995 #text_change 26-May-2000
C;Accession: S48789
R;Buerkle, X.Y.Z.; Frommer, W.B.
submitted to the EMBL Data Library, October 1994
A;Reference number: S48787
A;Accession: S48789
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-507 <BUE>
A;Cross-references: EMBL:X82276; NID:g575350; PID:g575351
C;Superfamily: common tobacco sucrose transport protein

Query Match 11.1%; Score 317.5; DB 2; Length 507;
Best Local Similarity 24.3%; Pred. No. 1.9e-16;
Matches 125; Conservative 87; Mismatches 203; Indels 99; Gaps 18;

```
Qy       4 RLWVSRLLRHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIGPVL 63
```

```
          :||          ::::|  :  |::   |  :: : | :  :|:  || : :   ||:
Db      23 KLW----------KIIMVASIAAGVQFGWALQLSLLTPYVQLLGIPHKFASFIWLCGPIS 72

Qy      64 GLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDP-----R 118
           |::  |::|  ||:   |:|||| || | :  : :::|||  |  | |    ||     :
Db      73 GMIVQPVVGYYSDNCSSRFGRRRGFIAAGAALVTIAVFLIGFAADL-GHATGDPLGKGSK 131

Qy     119 PLELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLL 178
           |  :|: ::|  :||    :   |  |||:||       | : : ::| :::|  |||
Db     132 PRAIAVFVVGFWILDVANNMLQGPCRALLADLSGGKARMRTSNAFFSFFMAVGNVLGYAA 191

Qy     179 PAID--------WDTSALAPYLGTQEECLF-GLLTLIFLTCVAATLLVAEEAALGPTEPA 229
           :            | |   |    : | |  :  |: || :| |  |  |  |   :
Db     192 GSYSRLCKIFPFSKTPACDIYCANLKSCFFIAVFLLLSLTILALT--VVRENELPEKDEH 249

Qy     230 EGLSAPSLSPHCCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFT 289
           |      :     :::: |   | :   |  |    || :  | :    :|:|   |
Db     250 E------IDEKAGARKSKVPF--FGEIFGALKDL----PRPMWILLLVTSLNWIARFPFF 297

Qy     290 LFYTDFVGEGLYQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRF 349
           |: ||:: : :| |  :   |    | || ||  |:||| |   :    || :: | ::
Db     298 LYDTDWMAKEVYGG--KVGDG----RLYDLGVHAGALGLLLNSVVLGFMSLSVEFLGKKI 351

Qy     350 GTRAVYLASVAAFPVAAGATCLSHSVAVV-------------TASAALTGFTFSALQI-- 394
           |     |  :  | :   | |:: :| |               |  |  :|    || :
Db     352 G-GVKRLWGILNFVL---AVCMALTVLVTKMAEKSRQYDAHGTLMAPTSGVKIGALTLFA 407

Qy     395 -----------LPYTLASLYHREK-------------QVFLPKYRGDTGGASSEDSLMTS 430
                      :|: |||::                : :|:      |   :|
Db     408 VLGIPLAVTFSVPFALASIFSSNAGSGQGLSLGVLNLAIVVPQMLVSIAGGPWDDLFGGG 467

Qy     431 FLPGPKPGAPFPNGHVGAGGSG-----LLPPPPA 459
           |||      |  | | |  ||     :|| |||
Db     468 NLPG------FIVGAVAAAASGILALTMLPSPPA 495


RESULT   9
S43142
sucrose transport protein - castor bean
N;Alternate names: sucrose carrier
C;Species: Ricinus communis (castor bean)
C;Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 26-May-2000
C;Accession: S43142
R;Weig, A.; Komor, E.
submitted to the EMBL Data Library, March 1994
A;Description: A sucrose carrier from Ricinus communis.
A;Reference number: S43142
A;Accession: S43142
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-533 <WEI>
A;Cross-references: EMBL:Z31561; NID:g468561; PID:g468562
C;Superfamily: common tobacco sucrose transport protein


  Query Match             10.9%;  Score 311;  DB 2;  Length 533;
  Best Local Similarity   23.3%;  Pred. No. 6.2e-16;
  Matches  134;  Conservative   95;  Mismatches  225;  Indels  120;  Gaps   19;

Qy      17 QLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIGPVLGLVCVPLLGSASD 76
           ::::|  :  |::   |  :: : | :  :|:   :   :   ||: |::  |::|  ||
Db      37 KVVMVASIAAGIQFGWALQLSLLTPYVQLLGIPHTWAAFIWLCGPISGMLVQPIVGYHSD 96

Qy      77 HWRGRYGRRRPFIWALSLGILLSLFLIPRA---GWLAG-LLCPDPRPLELALLILGVGLL 132
              |:|||||| : :  : :::|||  |   | |:|  |   |:  :|: ::|  :|
Db      97 RCTSRFGRRRPFIASGAAFVAIAVFLIGYAADLGHLSGDSLDKSPKTRAIAIFVVGFWIL 156

Qy     133 DFCGQVCFTPLEALLSDLF-RDPDHCRQAYSVYAFMISLGGCLGYLLPAID--------W 183
           |   :   |  |||:||       | | ::::| :::|  |||   |
Db     157 DVANNMLQGPCRALLADLSGTSQKKTRTANALFSFFMAVGNVLGYAAGAYTHLYKLFPFT 216

Qy     184 DTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGL--------SAP 235
           |:|   |    : | |  : |:    | |   | |:    |  | : :        |
Db     217 KTTACDVYCANLKSCFFISIVLLLSLTVLALSYVKEK----PWSPDQAVDNAEDDTASQA 272

Qy     236 SLSPHCCPCRARL--AFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYT 293
           | |    |    :  ||:||               | :  | :    :|:|   | || |
Db     273 SSSAQPMPFFGEILGAFKNL-------------KRPMWILLLVTCLNWIAWFPFLLFDT 318

Qy     294 DFVGEGLYQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRA 353
           |::|  :| |   :    |  : || ||| |:||| |   :    || :: | :  |
Db     319 DWMGREVYGG--DSSGSAEQLKLYDRGVRAGALGLMLNSVVLGFTSLGVEVLARGVG-GV 375

Qy     354 VYLASVAAFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPK 413
           |  :  | :   | ||: :| :||  |  |                          :
Db     376 KRLWGIVNFVL---AVCLAMTV-LVTKQAEST-------------------------RR 405

Qy     414 YRGDTGGASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVV 473
           :   :|||          :|                    ||||    |  |  :::  |
Db     406 FATVSGGAK---------VP--------------------LPPP---SGVKAGALALFAV 433

Qy     474 VGEPTEARV-VP----------GRGICLDLAILDSAFLLSQVAPSLFMGSIVQL--SQS 519
           :| |      :|           | |  | | :|: : :: |:  |:  |  |    :
Db     434 MGVPQAITYSIPFALASIFSNTSGAGQGLSLGVLNLSIVIPQMIVSVAAGPWDALFGGGN 493

Qy     520 VTAYMVSAAGLGLVAIYFATQVVFDKSDLAKYSA 553
           : |::| |       |:  | :   : |:    |
Db     494 LPAFVVGAVAALASGIFALTMLPSPQPDMPSAKA 527


RESULT  10
T14339
sucrose-proton transport protein - carrot
N;Alternate names: sucrose/H+ symporter protein
C;Species: Daucus carota (carrot)
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 21-Jul-2000
C;Accession: T14339
R;Shakya, R.; Sturm, A.
Plant Physiol. 118, 1473-1480, 1998
A;Title: Characterization of source- and sink-specific sucrose/H+ symporters from car
A;Reference number: Z17991; MUID:99063785
A;Accession: T14339
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-501 <SHA>
A;Cross-references: EMBL:Y16766; NID:g2969886; PIDN:CAA76367.1; PID:g2969887
A;Experimental source: cultivar Namtaise: leaf
C;Genetics:
A;Gene: SUT1a
C;Superfamily: common tobacco sucrose transport protein


  Query Match             10.7%;  Score 305.5;  DB 2;  Length 501;
  Best Local Similarity   21.6%;  Pred. No. 1.5e-15;
  Matches  124;  Conservative   91;  Mismatches  213;  Indels  145;  Gaps   14;

Qy      18 LLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIGPVLGLVCVPLLGSASDH 77
           || |  :  |::   |  :: : | : |:|:   : :::   ||: ||:  |::|  ||
Db      30 LLRVASVACGIQFGWALQLSLLTPYVQELGIPHAWSSIIWLCGPLSGLLVQPIVGHMSDQ 89

Qy      78 WRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLL--CPDPRPLELALLILGVGLLDFC 135
              :||||||| |    |:|:: :|  :  : |||    | : : :   ::|  :||
Db      90 CTSKYGRRRPFIVAGGTAIILAVIIIAHSADIGGLLGDTADNKTMAIVAFVIGFWILDVA 149

Qy     136 GQVCFTPLEALLSDLF-RDPDHCRQAYSVYAFMISLGGCLGYLLPAIDWDTSALAPYLGT 194
             :   |  |||:||    |    | | : ::  :::|  |||           |   | |
Db     150 NNMTQGPCRALLADLTGNDARRTRVANAYFSLFMAIGNVLGY----------ATGAYSGW 199

Qy     195 QEECLFGLLTLIFLTC-------------VAATLLVAEEAAL---------GPTEPAEGL 232
           :   | | :   : |              :  |  ::  ||           ||    :|
Db     200 YKVFPFSLTSSCTINCANLKSAFYIDIIFIIITTYISISAAKERPRISSQDGPQFSEDGT 259

Qy     233 SAPSLSPHCCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFY 292
           :   | |                | |  |      :| ::  : :    :|:    | ||
Db     260 AQ---SGHI-----------EEAFLWELFGTFRLLPGSVWVILLVTCLNWIGWFPFILFD 305
```

```
Qy     293 TDFVGEGLYQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTR 352
           ||::|  :| | |         : | :|||||: || :   :  : |::|::| : :|:
Db     306 TDWMGREIYGGEP------NQGQSYSDGVRMGAFGLMMNSVVLGITSVLMEKLCRIWGSG 359

Qy     353 AVYLASVAAFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLP 412
            ::                        |  |    | |: :: :   :: :
Db     360 FMW----------------------GLSNILMTICFFAMLLITFIAKNMDY-------- 388

Qy     413 KYRGDTGGASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRV 472
                                                 |  |||    |     : |
Db     389 ------------------------------------GTNPPPN---GIVISALIVFA 406

Qy     473 VVGEPTEARV-VP-----------GRGICLDLAILDSAFLLSQVAPSLFMGSIVQL---- 516
           ::| |      ||           | |  | : :|: | :: ||  ||  |   ||
Db     407 ILGIPLAITYSVPYALVSTRIESLGLGQGLSMGVLNLAIVVPQVIVSLGSGPWDQLFGGG 466

Qy     517 -SQSVTAYMVSAAGLGLVAIYFATQVVFDKSDL 548
            | :    :||   ||:|:    :   ||| |
Db     467 NSPAFVVAALSAFAAGLIALIAIRRPRVDKSRL 499
```

RESULT 11
S48788
sucrose transport protein - tomato (fragment)
C;Species: Lycopersicon esculentum (tomato)
C;Date: 27-Jan-1995 #sequence_revision 27-Jan-1995 #text_change 09-Jun-2000
C;Accession: S48788
R;Buerkle, X.Y.Z.; Frommer, W.B.
submitted to the EMBL Data Library, October 1994
A;Reference number: S48787
A;Accession: S48788
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-428 <BUE>
A;Cross-references: EMBL:X82275
C;Superfamily: common tobacco sucrose transport protein

Query Match 10.4%; Score 297.5; DB 2; Length 428;
Best Local Similarity 25.8%; Pred. No. 5.2e-15;
Matches 114; Conservative 82; Mismatches 163; Indels 83; Gaps 18;

```
Qy       4 RLWVSRLLRHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIGPVL 63
           :||           ::::|  :  |::   |  :: : | :  :|:  :| : :   ||:
Db      25 KLW----------KIIVVASIAAGVQFGWALQLSLLTPYVQLLGIPHRFASFIWLCGPIS 74

Qy      64 GLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDP-----R 118
           |::  |::|  ||:   |:|||||| | :  : :::|||  |  | |    ||     :
Db      75 GMIVQPVVGYYSDNCSSRFGRRRPFIAAGAALVTIAVFLIGFAADL-GHASGDPLGKGSK 133

Qy     119 PLELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPD-HCRQAYSVYAFMISLGGCLGY- 176
           |  :|: ::|  :||    :   |  |||:||        | | : ::|  ::|  |||
Db     134 PRAIAVFVVGFWILDVANNMLQGPCRALLADLSGGKSGKMRTANAFFSFFKAVGNILGYA 193

Qy     177 ------LLPAIDWD-TSALAPYLGTQEECLF-GLLTLIFLTCVAATLL----VAEEAALG 224
                |     : | |   |    : | |  :  |: || :| ||:    : |:  |
Db     194 AGSYSRLFKVFPFSKTKACDMYRANLKSCFFIAIFLLLSLTTLALTLVRENELPEKEELE 253

Qy     225 PTEPAEGLSAPSLSPHCCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMA 284
             |   |             :::: |   | :   |  |    || :  | :    :|:|
Db     254 IDEKLSGAG-----------KSKVPF--FGEIFGALKDL----PRPMWILLLVTCLNWIA 296

Qy     285 LMTFTLFYTDFVGEGLYQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDR 344
              | |: ||:: : :: |    : |    : || ||| |:||| ||  :    || ::
Db     297 WFPFFLYDTDWMAKEVFGG----QVGD--AKLYDLGVRAGALGLLLQSVVLGFMSLGVEF 350

Qy     345 LVQRFGTRAVYLASVAAFPVAAGATCLSHSVAVV---------TASAALTGFT------- 388
           | :: |  |  |  :  | :   | ||: :: |           |:  | | |
Db     351 LGKKIG-GAKRLWGILNFVL---AICLAMTILVTKMAEKSRGHDAAGTLMGPTPGVKIGA 406

Qy     389 ---FSALQI-------LPYTLA 400
              |:|| |       :|: ||
Db     407 LLLFAALGIPLAVTFSIPFALA 428
```

RESULT 12
S38657
sucrose transport protein ptp1 - common plantain
N;Alternate names: sucrose transporter ptp1
C;Species: Plantago major (common plantain)
C;Date: 19-Mar-1997 #sequence_revision 06-Jun-1997 #text_change 26-May-2000
C;Accession: S38657
R;Gahrtz, M.; Stolz, J.; Sauer, N.
submitted to the EMBL Data Library, November 1993
A;Reference number: S38657
A;Accession: S38657
A;Molecule type: mRNA
A;Residues: 1-510 <GAH>
A;Cross-references: EMBL:X75764; NID:g415987; PIDN:CAA53390.1; PID:g415988
C;Genetics:
A;Gene: ptp1
C;Superfamily: common tobacco sucrose transport protein
C;Keywords: sugar transport

Query Match 10.3%; Score 294; DB 2; Length 510;
Best Local Similarity 24.7%; Pred. No. 1.1e-14;
Matches 126; Conservative 77; Mismatches 199; Indels 108; Gaps 19;

```
Qy      18 LLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIGPVLGLVCVPLLGSASDH 77
           : ||  :  |::   |  :: : | :  :|:  |: : :   ||: |::  |::|  ||:
Db      29 IFLVAAIAAGVQFGWALQLSLLTPYVQLLGIPHKWASYIWLCGPISGMIVQPVVGYFSDN 88

Qy      78 WRGRYGRRRPFIWALSLGIL-LSLFLIPRAGWLAGL-------LCPDPRPLELALLILGV 129
              |:|||||| |   |:: ::: ||   |: | |       |    :|  : : : |
Db      89 CTSRFGRRRPFI-AAGAGLVGVAVVLI---GFAADLGHAGGDSLGDGLKPRAIGVFVFGF 144

Qy     130 GLLDFCGQVCFTPLEALLSDLF-RDPDHCRQAYSVYAFMISLGGCLGYLLPAID------ 182
            :||    :   |  |||:||   :       | | ::| :::|  |||   :
Db     145 WILDVANNMLQGPCRALLADLSGGNTKKMANANSFFSFFMAVGNVLGYAAGSYSRMYKVF 204

Qy     183 --WDTSALAPYLGTQEEC-LFGLLTLIFLTCVAATLL-----VAEEAALGPTEPAEGLSA 234
               |  |    |    : | :  :  || || :| :::     ||||     |   :|
Db     205 PFSKTKACDIYCANLKSCFIISITLLITLTTLALSIVREKRHVAEEQV---TAAKKGFKI 261

Qy     235 PSLSPHCCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTD 294
           |                     : | |      :|| :  | :    :|:|   | || ||
Db     262 P-------------------VFPELFGALKDLPRPMWVLLLVTALNWIAWFGFLLFDTD 301

Qy     295 FVGEGLYQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFG---- 350
           ::|  :| | :    |    |::||  |:||| |   :    || :  : :  |
Db     302 WMGREVY-GETQQHKAPELAVIYNKGVSAGALGLMLNSIVLGFASLGVQYMARALGGVKR 360

Qy     351 -------TRAVYLA-SVAAFPVAAGATCLSHSVAVVTASAALTG--FTFSALQI------ 394
                  |: |  ::    ||:     |: |     |:  |     ||||
Db     361 LWGVVNFILAICLCMTIVITKVASHHRPYSNGVLQTPESSVKIGALVVFSALGIPLAITF 420

Qy     395 -LPYTLASLYHREKQVFLPKYRGDTGGASSEDSL-----------MTSFLPGPKPG--- 438
           :|: |||:|               | |:    ||            : |   ||
Db     421 SVPFALASIY-----------STTTGSGQGLSLGVLNLAIVIPQMIVSVASGPWDAMFG 468

Qy     439 ----APFPNGHVGAGGSG-----LLPPPPA 459
                |  | | |  ||      :|| |||
Db     469 GGNLPAFVVGAVAAAASGIFAFTMLPSPPA 498
```

RESULT 13
S51114
sucrose-proton symporter - beet
C;Species: Beta vulgaris (beet)
C;Date: 07-May-1995 #sequence_revision 01-Sep-1995 #text_change 26-May-2000
C;Accession: S51114
R;Westram, A.; Eckhardt, U.; Frommer, W.B.; Riesmeier, J.W.
submitted to the EMBL Data Library, January 1995
A;Description: Sequence of a sugar beet sucrose transporter cDNA.
A;Reference number: S51114

A;Accession: S51114
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-523 <WES>
A;Cross-references: EMBL:X83850; NID:g633171; PID:g633172
C;Superfamily: common tobacco sucrose transport protein

Query Match 10.3%; Score 294; DB 2; Length 523;
Best Local Similarity 24.3%; Pred. No. 1.2e-14;
Matches 125; Conservative 78; Mismatches 199; Indels 112; Gaps 20;

```
Qy      17 QLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIGPVLGLVCVPLLGSASD 76
           :| ||  :  |::   |  :: : | :  :|:   :   :   ||: |::  | :|  ||
Db      38 KLALVASIAAGVQFGWALQLSLLTPYVQLLGIPHTWAPYIWLCGPISGMIVQPTVGYYSD 97

Qy      77 HWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLA--GLLCPD-----PRPLELALLILGV 129
             ::||| |||   ::|  |  | :   |: |  |    |     |:|  :|: ::|
Db      98 RCTSKFGRRSPFI---AVGATLVGFAVSLIGFAADIGHATGDPNGNVPKPRAIAVFVVGF 154

Qy     130 GLLDFCGQVCFTPLEALLSDLFRDPD-HCRQAYSVYAFMISLGGCLGY---------LLP 179
           :||        |  |||:|:        | | : ::| ::||   ||         : |
Db     155 WILDVANNTLQGPCRALLADMAAGSQAKTRYANAFFSFFMALGNIGGYAGSYGRLYTVFP 214

Qy     180 AIDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSP 239
             | |   |    : | |  :||: :  :  : |  :| |     |    |     :|
Db     215 FT--HTKACDTYCANLKSCFFISITLLIVLTILALSVVRER----PFTLDEIQEEENLKN 268

Qy     240 HCCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEG 299
           :   | ||| |   | |   |  |    |: :  | :    :|:|   | || ||::|:
Db     269 NTGGC-ARLPF--FGQLFGALKDL----PKPMLILLLVTCLNWIAWFPFLLFDTDWMGKE 321

Qy     300 LYQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFG--TRAVYLA 357
           :| |          : || ||  |:||| :   :  : || :::| :  |   |   :
Db     322 VYGGT------VGEGKAYDMGVHAGALGLMINSVVLGIMSLGIEKLARLVGGVKRLWGIV 375

Qy     358 SVAAFPVAAGATCLSHSVAVVTASA----------------------------ALTGF 387
           ::        | ||: :: :|| ||                             |: |
Db     376 NLIL------AVCLAMTI-LVTKSAEHYRATHHVPGAIGPPLPPPGVKGGALAIFAVLGI 428

Qy     388 TFSALQILPYTLASLYHREK-----------------QVFLPKYRGDTGGASSEDSLMT 429
             :    :|: |||::                      |:|:    |         |:|
Db     429 PLAITFSIPFALASIFSASSGSGQGLSLGVLNLAIVVPQMFVSVTSGPW------DAL-- 480

Qy     430 SFLPGPKPGAPFPNGHVGAGGSG-----LLPPPP 458
             |  |  |   |  | | |  |      ||||||
Db     481 -FGGGNLPA--FVVGAVAATASAILSFTLLPPPP 511
```

RESULT 14
F96741
probable sucrose transport protein F17M19.4 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 23-Mar-2001
C;Accession: F96741
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719
A;Accession: F96741
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-512 <STO>
A;Cross-references: GB:AE005173; NID:g6978914; PIDN:AAF34306.1; GSPDB:GN00141
C;Genetics:
A;Gene: F17M19.4
A;Map position: 1
C;Superfamily: common tobacco sucrose transport protein

Query Match 10.2%; Score 292.5; DB 2; Length 512;
Best Local Similarity 23.8%; Pred. No. 1.5e-14;
Matches 120; Conservative 87; Mismatches 211; Indels 87; Gaps 18;

```
Qy      17 QLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIGPVLGLVCVPLLGSASD 76
           ::: |  :  |::   |  :: : | :  :|:  |: : :   ||: |::  |::|  ||
Db      33 KIISVASIAAGVQFGWALQLSLLTPYIQLLGIPHKWSSYMWLCGPISGMIVQPIVGYHSD 92

Qy      77 HWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGL-------LCPDPRPLELALLILGV 129
             |:|||||| |    : :|:|||   |: | :        |    |   : : : |
Db      93 RCESRFGRRRPFIAAGVALVAVSVFLI---GFAADMGHSFGDKLENKVRTRAIIIFLTGF 149

Qy     130 GLLDFCGQVCFTPLEALLSDLFR-DPDHCRQAYSVYAFMISLGGCLGY----------LL 178
           ||         |  | |:||   |    | | : ::| :::|  |||         :
Db     150 WFLDVANNTLQGPCRAFLADLAAGDAKKTRVANACFSFFMAVGNVLGYAAGSYTNLHKMF 209

Qy     179 PAIDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLS 238
           |     | |   |    : | |  :||: :   ::   | ::       | :|      |
Db     210 PFT--MTKACDIYCANLKTCFFLSITLLLIVTFSSLWYVKDK----QWSPPQGDKEEKTS 263

Qy     239 PHCCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGE 298
             | |   | :   :       | | :  | :  : :|:|   | |: ||::|
Db     264 --------SLFF--FGEIFGAVR----HMKRPMVMLLIVTVINWIAWFPFILYDTDWMGR 309

Qy     299 GLYQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLAS 358
           :| |   ::   ::: ||:||: |:|||     :    || :: : :: |  |  |
Db     310 EVYGG--NSDGDERSKKLYDQGVQAGALGLMFNSILLGFVSLGVESIGRKMG-GAKRLWG 366

Qy     359 VAAFPVAAGATCLSHSVAVVTA--------------SAALTGFTFSALQIL--------- 395
             | :| |   |: :| |  :                |: :    ||   :|
Db     367 CVNFILAIG---LAMTVLVTKSAEHHREIAGPLAGPSSGIKAGVFSLFTVLGIPLAITYS 423

Qy     396 -PYTLASLYHREKQVFLPKYRGDTGGASSEDSLMTSFLPGPKP----GAPFPN---GHVG 447
            |: |||::             |    |     :: ||  ||      |   |:  | :
Db     424 IPFALASIFSTNSGAGQGLSLGVLNIAICIPQMIVSFSSGPLDAQFGGGNLPSFVVGAIA 483

Qy     448 AGGSGLL-------PPP--PALCGA 463
           |  ||:|       |||  ||: ||
Db     484 AAVSGVLALTVLPSPPPDAPAMSGA 508
```

RESULT 15
T02982
probable sucrose transport protein - rice
C;Species: Oryza sativa (rice)
C;Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 21-Jul-2000
C;Accession: T02982
R;Hirose, T.; Imaizumi, N.; Scofield, G.N.; Furbank, R.T.; Ohsugi, R.
Plant Cell Physiol. 38, 1389-1396, 1997
A;Title: cDNA cloning and tissue specific expression of a gene for sucrose transporte
A;Reference number: Z14809; MUID:98182940
A;Accession: T02982
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-537 <HIR>
A;Cross-references: EMBL:D87819; NID:g2723470; PIDN:BAA24071.1; PID:g2723471
A;Experimental source: cultivar Nipponbare, leaf
C;Genetics:
A;Gene: SUT1
C;Superfamily: common tobacco sucrose transport protein

Query Match 10.2%; Score 292; DB 2; Length 537;
Best Local Similarity 26.7%; Pred. No. 1.7e-14;
Matches 115; Conservative 72; Mismatches 167; Indels 76; Gaps 19;

```
Qy      17 QLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIGPVLGLVCVPLLGSASD 76
           :|:|  ::  |::   |  :: : | :  :|:     : :   ||: |:|  | :|  ||
Db      51 RLILSGMVAGGVQYGWALQLSLLTPYVQTLGLSHALTSFMWLCGPIAGMVVQPCVGLYSD 110
```

```
Qy      77 HWRGRYGRRRPFIWALSLGILLSLFLI---PRAGWLAGLLCPD------PRPLELALLIL 127
Db     111 RCTSKWGRRRPYILTGCVLICLAVVVIGFSADIGYAMGDTKEDCSVYHGSRWHAAIVYVL 170

Qy     128 GVGLLDFCGQVCFTPLEALLSDLF--RDPDHCRQAYSVYAFMISLGGCLGYLLPAID--- 182
Db     171 GFWLLDFSNNTVQGPARALMADLSGRHGPG---TANSIFCSWMAMGNILGYSSGSTNNWH 227

Qy     183 -WDTSALAPYLGTQEEC-----LFG--LLTLIFLT-CVAATLLVAEE------AALGPT- 226
Db     228 KW-----FPFLKTRACCEACANLKGAFLVAVIFLSLCLVITLIFAKEVPFKGNAAL-PTK 281

Qy     227 --EPAEGLSAPSLSPHCCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMA 284
Db     282 SNEPAEPEGTG-------PLAVLKGFRNLPTGMPSV------------LIVTGL-TWLS 320

Qy     285 LMTFTLFYTDFVGEGLYQGVPR-AEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMD 343
Db     321 WFPFILYDTDWMGREIYHGDPKGTDPQIEA---FNQGVRAGAFGLLLNSIVLGFSSFLIE 377

Qy     344 RLVQRFGTRAVYLASVAAFPVAAGATCL--------SHSVA--VVTASAALTGFTFSALQ 393
Db     378 PMCRKVGPRVVWVTSNFLVCIAMAATALISFWSLKDFHGTVQKAITADKSIKAVCLVLFA 437

Qy     394 ILPYTLASLY 403
Db     438 FLGVPLAVLY 447
```

Search completed: June 26, 2002, 13:45:46
Job time: 3939 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: June 26, 2002, 13:49:02 ; Search time 30.25 Seconds
(without alignments)
102.812 Million cell updates/sec

Title: US-09-838-785-25
Perfect score: 28
Sequence: 1 DFVGEGLYQGVPRAEGTEARRHYDEGVR 28

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 747574 seqs, 111073796 residues

Word size : 0

Total number of hits satisfying chosen parameters: 268091

Minimum DB seq length: 0
Maximum DB seq length: 28

Post-processing: Listing first 45 summaries

Database : A_Geneseq_032802:*
1: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	13	46.4	25	22	AAU69815	Human prostate pro
2	13	46.4	25	22	AAM01170	Human prostate-spe
3	13	46.4	25	22	AAG99055	Human prostate-spe
4	13	46.4	27	20	AAW85071	PS108 protein deri
5	13	46.4	27	21	AAB29271	Human prostate-rel
6	13	46.4	27	22	AAU69941	Human prostate pro
7	13	46.4	27	22	AAU04208	Prostate-specific
8	13	46.4	27	22	AAM01296	P501S peptide 296-
9	11	39.3	15	22	AAU69813	Human prostate pro
10	11	39.3	15	22	AAM01168	Human prostate-spe
11	11	39.3	15	22	AAG99053	Human prostate-spe
12	6	21.4	17	22	AAU69814	Human prostate pro
13	6	21.4	17	22	AAM01169	Human prostate-spe
14	6	21.4	17	22	AAG99054	Human prostate-spe
15	5	17.9	8	22	AAE01151	Peptide #1 of pepp
16	5	17.9	9	19	AAY21072	Human p53 cellular
17	5	17.9	9	20	AAW92745	M. spicata geranyl
18	5	17.9	9	22	AAG88323	HER2/NEU DR superm
19	5	17.9	10	22	AAG95547	Human complementar
20	5	17.9	13	17	AAR94992	Nsk2 receptor C-te
21	5	17.9	14	17	AAR89428	KLH-conjugated pep
22	5	17.9	14	18	AAW15066	Bicyclic neurotrop
23	5	17.9	14	19	AAW65951	Peptide derived fr
24	5	17.9	15	12	AAR15201	C-erb B-2 epitope.
25	5	17.9	15	14	AAR37466	c-erbB2 hydrophobi
26	5	17.9	15	19	AAY21185	Human bcl2 proto-o
27	5	17.9	15	22	AAG88324	HER2/NEU DR superm
28	5	17.9	17	15	AAR57110	Human megakaryocyt
29	5	17.9	18	16	AAR79980	Signal transductio
30	5	17.9	18	21	AAY52610	v-myb encoded onco
31	5	17.9	19	18	AAW30142	Tandem SH2 ligand
32	5	17.9	19	20	AAY41954	Rheumatoid arthrit
33	5	17.9	20	16	AAR79992	ARH1 peptide mimet
34	5	17.9	21	17	AAR84090	Nsk2 receptor alte
35	5	17.9	21	19	AAW62571	Nsk2 alternatively
36	5	17.9	22	19	AAW79283	Tomato ringspot vi
37	5	17.9	22	22	AAY72662	Human headpin (for
38	5	17.9	26	21	AAY52314	T-cell receptor ze
39	5	17.9	26	21	AAY52333	Phosphorylated ITA
40	5	17.9	26	21	AAY52334	Phosphorylated ITA
41	5	17.9	26	21	AAY52335	Phosphorylated ITA
42	5	17.9	27	22	AAB98733	Human TCR zeta cha
43	5	17.9	27	22	AAB98780	Human TCR zeta cha
44	4	14.3	4	18	AAW19148	Isoelectric point
45	4	14.3	5	9	AAP82551	Anti-inflammatory

ALIGNMENTS

RESULT 1
AAU69815
ID AAU69815 standard; Peptide; 25 AA.
XX
AC AAU69815;
XX
DT 30-JAN-2002 (first entry)
XX
DE Human prostate protein P703P antigenic epitope #12.
XX
KW Human; prostate cancer; ss; cytostatic; immunostimulant; tumour;
KW antigen; epitope.
XX
OS Homo sapiens.
XX
PN WO200173032-A2.
XX
PD 04-OCT-2001.
XX
PF 27-MAR-2001; 2001WO-US09919.
XX
PR 27-MAR-2000; 2000US-0536857.
PR 09-MAY-2000; 2000US-0568100.
PR 12-MAY-2000; 2000US-0570737.
PR 13-JUN-2000; 2000US-0593793.
PR 27-JUN-2000; 2000US-0605783.
PR 10-AUG-2000; 2000US-0636215.
PR 29-AUG-2000; 2000US-0651236.
PR 06-SEP-2000; 2000US-0657279.
PR 02-OCT-2000; 2000US-0679426.
PR 10-OCT-2000; 2000US-0685166.
XX
PA (CORI-) CORIXA CORP.
XX

PI Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Kalos MD;
PI Fanger GR, Retter MW, Stolk JA, Day CH, Vedvick TS, Carter D;
PI Li SX, Wang A, Skeiky YAW, Hepler WT, Henderson RA;
XX
DR WPI; 2001-639232/73.
XX
PT New human prostate-specific polypeptides and polynucleotides useful for
PT the diagnosis and treatment of cancer, especially prostate cancer -
XX
PS Claim 2; Page 403; 579pp; English.
XX
CC The invention relates to isolated prostate-specific
CC polynucleotides, polypeptides, fusion proteins of the polypeptides,
CC antibodies raised against the polypeptides (or antigenic epitopes
CC derived from them) and antigen-presenting cells expressing the
CC polypeptides. The antibodies are useful for detecting the presence of
CC cancer, especially prostate cancer. The polypeptides, polynucleotides and
CC the antigen-presenting cells are useful for stimulating and/or expanding
CC T cells specific for a tumour protein, and for inhibiting the development
CC of cancer especially prostate cancer. Compositions comprising the
CC polynucleotide and/or polypeptide are useful for stimulating an immune
CC response, and for treating cancer. The oligonucleotide is useful for
CC detecting cancer. The present sequence is a prostate specific
CC peptide of the invention. The peptides either represent antigenic
CC epitopes or domains of prostate specific proteins.
XX
SQ Sequence 25 AA;

Query Match 46.4%; Score 13; DB 22; Length 25;
Best Local Similarity 100.0%; Pred. No. 6e-06;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 VGEGLYQGVPRAE 15
||||||||||||
Db 1 vgeglyqgvprae 13

RESULT 2
AAM01170
ID AAM01170 standard; Peptide; 25 AA.
XX
AC AAM01170;
XX
DT 04-OCT-2001 (first entry)
XX
DE Human prostate-specific peptide used in epitope mapping studies #27.
XX
KW Human; prostate cancer; prostate-specific; diagnosis; vaccine;
KW cytostatic; gene therapy; metastasis.
XX
OS Homo sapiens.
XX
PN WO200151633-A2.
XX
PD 19-JUL-2001.
XX
PF 16-JAN-2001; 2001WO-US01574.
XX
PR 14-JAN-2000; 2000US-0483672.
XX
PA (CORI-) CORIXA CORP.
XX
PI Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Reed SG;
PI Kalos MD, Fanger GR, Day CH, Retter MW, Stolk JA, Skeiky YAW;
PI Wang A, Meagher MJ;
XX
DR WPI; 2001-425873/45.
XX
PT New polynucleotide encoding a prostate-specific protein, for
PT diagnosing, monitoring and treating prostate cancer in a patient and
PT for use in vaccines -
XX
PS Claim 2; Page 403; 543pp; English.
XX
CC The present invention describes polynucleotide sequences (I) which encode
CC prostate-specific proteins (II). (I) and (II) have cytostatic activity,
CC and can be used in vaccine production and gene therapy. (I), (II),
CC antibodies to (II), fusion proteins comprising (II), and isolated
CC T cells prepared using (I) or (II) are used treat cancer in a patient.
CC (I) and the antibodies are also used in the detection of cancer in a
CC patient. The cancer that is diagnosed or treated is particularly
CC prostate cancer. (I) and (II) can be used in vaccines. The antibodies or
CC (I) can be used for monitoring the progression of cancer in a patient.
CC (I) and (II) can also be used to improve diagnostic and therapeutic
CC methods for prostate cancer. They can indicate the level of metastasis
CC as well as the prostate volume. AAH93357 to AAH93944 and AAM01115 to
CC AAM01318 represent polynucleotide and amino acid sequences used in the
CC exemplification of the present invention.
XX
SQ Sequence 25 AA;

Query Match 46.4%; Score 13; DB 22; Length 25;
Best Local Similarity 100.0%; Pred. No. 6e-06;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 VGEGLYQGVPRAE 15
||||||||||||
Db 1 vgeglyqgvprae 13

RESULT 3
AAG99055
ID AAG99055 standard; Peptide; 25 AA.
XX
AC AAG99055;
XX
DT 25-SEP-2001 (first entry)
XX
DE Human prostate-specific epitope mapping peptide #27.
XX
KW Human; prostate cancer; therapy; diagnosis; cat eye syndrome;
KW chromosome 22q11.2; prostate-specific protein; chromosome 1;
KW prostate specific antigen; PSA.
XX
OS Homo sapiens.
XX
PN WO200134802-A2.
XX
PD 17-MAY-2001.
XX
PF 09-NOV-2000; 2000WO-US30904.
XX
PR 12-NOV-1999; 99US-0439313.
PR 18-NOV-1999; 99US-0443686.
XX
PA (CORI-) CORIXA CORP.
XX
PI Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Reed SG;
PI Kalos MD, Retter MW, Stolk JA, Day CH, Skeiky YAW, Wang A;
XX
DR WPI; 2001-308785/32.
XX
PT Isolated polypeptide comprising at least an immunogenic portion of a
PT prostate-specific protein, useful in the diagnosis and therapy of
PT prostate cancer -
XX
PS Claim 3; Page 303; 325pp; English.
XX
CC The present invention describes an isolated polypeptide (P1) comprising
CC at least an immunogenic portion of a prostate-specific protein, or its
CC variant. Also described are polynucleotides (N1) encoding (P1). (P1) and
CC (N1) have cytostatic activity and can be used in vaccine production.
CC The polypeptides, nucleic acids and antibodies from the present
CC invention are useful in the diagnosis and therapy of prostate cancer.

CC Prostate specific genes P704P, P712P, P774P, P775P and B305D are located
CC in a genomic region on chromosome 22q11.2 known as the Cat Eye Syndrome
CC region. Prostate specific antigen (PSA) P501S was located on
CC chromosome 1. AAH84671 to AAH85143 and AAG99000 to AAG99077 represent
CC polynucleotide and polypeptide sequences used in the exemplification
CC of the present invention.
XX
SQ Sequence 25 AA;

Query Match 46.4%; Score 13; DB 22; Length 25;
Best Local Similarity 100.0%; Pred. No. 6e-06;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 VGEGLYQGVPRAE 15
||||||||||||
Db 1 vgeglyqgvprae 13

RESULT 4
AAW85071
ID AAW85071 standard; Peptide; 27 AA.
XX
AC AAW85071;
XX
DT 12-FEB-1999 (first entry)
XX
DE PS108 protein derived peptide.
XX
KW PS108 gene; prostate disease; benign prostatic hyperplasia; BPH;
KW prostatitis; prostatic intraepithelial neoplasia; PIN; cancer;
KW drug screening; gene therapy.
XX
OS Homo sapiens.
XX
PN WO9850567-A1.
XX
PD 12-NOV-1998.
XX
PF 01-MAY-1998; 98WO-US08930.
XX
PR 02-MAY-1997; 97US-0850713.
XX
PA (ABBO) ABBOTT LAB.
XX
PI Billing-Medel PA, Cohen M, Colpitts TL, Friedman PN;
PI Gordon J, Granados EN, Hodges SC, Klass MR, Kratochvil JD;
PI Roberts-Rapp L, Russell JC, Stroupe SD;
XX
DR WPI; 1999-034731/03.
XX
PT New isolated prostate-specific polynucleotides - used to develop
PT products for the diagnosis and treatment of prostate diseases, e.g.
PT benign hyperplasia, prostatic or prostate cancer
XX
PS Claim 17; Page 101; 122pp; English.
XX
CC The present sequence is derived from the PS108 protein. The
CC specification describes a method for detecting the presence
CC of a target PS108 polynucleotide in a test sample. The method
CC comprises contacting the test sample with at least 1 PS108-specific
CC polynucleotide or complement, and detecting the presence of the target
CC PS108 polynucleotide. The products can be used for detecting, diagnosing,
CC staging, monitoring, prognosticating, in vivo imaging, preventing or
CC treating, or determining predisposition to diseases or conditions of
CC the prostate such as benign prostatic hyperplasia (BPH), prostatitis,
CC prostatic intraepithelial neoplasia (PIN) and cancer. In particular the
CC products can be used in drug screening and gene therapy.
XX
SQ Sequence 27 AA;

Query Match 46.4%; Score 13; DB 20; Length 27;
Best Local Similarity 100.0%; Pred. No. 6.4e-06;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 16 GTEARRHYDEGVR 28
||||||||||||
Db 12 gtearrhydegvr 24

RESULT 5
AAB29271
ID AAB29271 standard; Protein; 27 AA.
XX
AC AAB29271;
XX
DT 07-FEB-2001 (first entry)
XX
DE Human prostate-related PS108 partial protein sequence SEQ ID NO: 39.
XX
KW Human; prostate cancer; PS108; antibody; tumour; metastasis.
XX
OS Homo sapiens.
XX
PN US6130043-A.
XX
PD 10-OCT-2000.
XX
PF 01-MAY-1998; 98US-0071710.
XX
PR 02-MAY-1997; 97US-0850713.
XX
PA (ABBO) ABBOTT LAB.
XX
PI Cohen M, Colpitts TL, Friedman PN, Gordon J, Granados EN;
PI Billing-Medel PA, Klass MR, Roberts-Rapp L, Stroupe SD, Yu H;
PI Kratochvil JD, Russell JC, Hodges SC;
XX
DR WPI; 2000-655655/63.
XX
PT Methods for detecting target prostate-specific polynucleotides or
PT diseases of the prostate (e.g. prostate cancer), comprising detecting
PT the presence of any of PS108 nucleic acid sequences in a test sample -
XX
PS Example 10; Column 87-88; 55pp; English.
XX
CC The present invention is related to a number of partial coding and
CC protein sequences for the human prostate tissue protein PS108. These
CC sequences can be used in the diagnosis and prognosis of prostate
CC diseases, particularly prostate cancer. They can also be used to produce
CC antibodies which can be used in treatment. The present sequence is one
CC of the PS108 partial protein sequences.
XX
SQ Sequence 27 AA;

Query Match 46.4%; Score 13; DB 21; Length 27;
Best Local Similarity 100.0%; Pred. No. 6.4e-06;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 16 GTEARRHYDEGVR 28
||||||||||||
Db 12 gtearrhydegvr 24

RESULT 6
AAU69941
ID AAU69941 standard; Peptide; 27 AA.
XX
AC AAU69941;
XX
DT 30-JAN-2002 (first entry)
XX
DE Human prostate protein P501S peptide 296-322.
XX

```
KW   Human; prostate cancer; ss; cytostatic; immunostimulant; tumour;
KW   antigen; epitope.
XX
OS   Homo sapiens.
XX
PN   WO200173032-A2.
XX
PD   04-OCT-2001.
XX
PF   27-MAR-2001; 2001WO-US09919.
XX
PR   27-MAR-2000; 2000US-0536857.
PR   09-MAY-2000; 2000US-0568100.
PR   12-MAY-2000; 2000US-0570737.
PR   13-JUN-2000; 2000US-0593793.
PR   27-JUN-2000; 2000US-0605783.
PR   10-AUG-2000; 2000US-0636215.
PR   29-AUG-2000; 2000US-0651236.
PR   06-SEP-2000; 2000US-0657279.
PR   02-OCT-2000; 2000US-0679426.
PR   10-OCT-2000; 2000US-0685166.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Kalos MD;
PI   Fanger GR, Retter MW, Stolk JA, Day CH, Vedvick TS, Carter D;
PI   Li SX, Wang A, Skeiky YAW, Hepler WT, Henderson RA;
XX
DR   WPI; 2001-639232/73.
XX
PT   New human prostate-specific polypeptides and polynucleotides useful for
PT   the diagnosis and treatment of cancer, especially prostate cancer -
XX
PS   Claim 2; Page 558; 579pp; English.
XX
CC   The invention relates to isolated prostate-specific
CC   polynucleotides, polypeptides, fusion proteins of the polypeptides,
CC   antibodies raised against the polypeptides (or antigenic epitopes
CC   derived from them) and antigen-presenting cells expressing the
CC   polypeptides. The antibodies are useful for detecting the presence of
CC   cancer, especially prostate cancer. The polypeptides, polynucleotides and
CC   the antigen-presenting cells are useful for stimulating and/or expanding
CC   T cells specific for a tumour protein, and for inhibiting the development
CC   of cancer especially prostate cancer. Compositions comprising the
CC   polynucleotide and/or polypeptide are useful for stimulating an immune
CC   response, and for treating cancer. The oligonucleotide is useful for
CC   detecting cancer. The present sequence is a prostate specific
CC   peptide of the invention. The peptides either represent antigenic
CC   epitopes or domains of prostate specific proteins.
XX
SQ   Sequence   27 AA;


  Query Match             46.4%;  Score 13;  DB 22;  Length 27;
  Best Local Similarity   100.0%;  Pred. No. 6.4e-06;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 VGEGLYQGVPRAE 15
           |||||||||||||
Db       1 vgeglyqgvprae 13


RESULT   7
AAU04208
ID   AAU04208 standard; Peptide; 27 AA.
XX
AC   AAU04208;
XX
DT   23-OCT-2001  (first entry)
XX
DE   Prostate-specific polypeptide PS108 #4.
XX
KW   Prostate; PS108; immunogen; drug screening; image localisation;
KW   diagnostic; therapeutic; prostate tissue disease; cancer; metastasis;
KW   expressed sequence tag; EST.
XX
OS   Homo sapiens.
XX
PN   US6252047-B1.
XX
PD   26-JUN-2001.
XX
PF   15-MAR-2000; 2000US-0525397.
XX
PR   01-MAY-1998;   98US-0071710.
PR   02-MAY-1997;   97US-0850713.
XX
PA   (ABBO ) ABBOTT LAB.
XX
PI   Billing-Medel PA, Cohen M, Colpitts TL, Friedman PN, Gordon J;
PI   Granados EN, Hodges SC, Klass MR, Kratochvil JD, Roberts-Rapp L;
PI   Russell JC, Stroupe SD, Yu H;
XX
DR   WPI; 2001-424488/45.
XX
PT   Novel PS108 polypeptide useful in assays for detecting antibodies to
PT   prostate tissue, and as immunogens to produce PS108 antibodies -
XX
PS   Claim 1; Column 87; 55pp; English.
XX
CC   The sequence represents the amino acid sequence of prostate-specific
CC   PS108 #4. The PS108 polypeptide is useful in assays for detecting
CC   antibodies to prostate tissue, and as an immunogen to produce antibodies.
CC   The polypeptide is useful for screening compounds which specifically bind
CC   to the polypeptide and for screening for drugs, compounds, or any other
CC   agent which can be used to treat diseases associated with PS108. The
CC   antibody is useful to detect, or for image localisation of, PS108 antigen
CC   in a patient, for detecting or diagnosing a disease or condition, as
CC   delivery agents for therapeutic agents as well as for diagnostic tests
CC   and for screening for diseases or conditions associated with PS108,
CC   especially cancer. The antibody is also useful for generating chimeric
CC   antibodies for therapeutic use, for inhibiting the biological activity of
CC   PS108, in therapy (for e.g. to treat prostate tissue disease including
CC   prostate cancer and its metastases), and to detect the presence of any
CC   polypeptide in a test sample which shares one or more antigenic
CC   determinants with the PS108 polypeptide.
XX
SQ   Sequence   27 AA;


  Query Match             46.4%;  Score 13;  DB 22;  Length 27;
  Best Local Similarity   100.0%;  Pred. No. 6.4e-06;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      16 GTEARRHYDEGVR 28
           |||||||||||||
Db      12 gtearrhydegvr 24


RESULT   8
AAM01296
ID   AAM01296 standard; Peptide; 27 AA.
XX
AC   AAM01296;
XX
DT   04-OCT-2001  (first entry)
XX
DE   P501S peptide 296-322.
XX
KW   Human; prostate cancer; prostate-specific; diagnosis; vaccine;
KW   cytostatic; gene therapy; metastasis.
XX
OS   Homo sapiens.
XX
PN   WO200151633-A2.
XX
```

PD 19-JUL-2001.
XX
PF 16-JAN-2001; 2001WO-US01574.
XX
PR 14-JAN-2000; 2000US-0483672.
XX
PA (CORI-) CORIXA CORP.
XX
PI Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Reed SG;
PI Kalos MD, Fanger GR, Day CH, Retter MW, Stolk JA, Skeiky YAW;
PI Wang A, Meagher MJ;
XX
DR WPI; 2001-425873/45.
XX
PT New polynucleotide encoding a prostate-specific protein, for
PT diagnosing, monitoring and treating prostate cancer in a patient and
PT for use in vaccines -
XX
PS Claim 2; Page 522; 543pp; English.
XX
CC The present invention describes polynucleotide sequences (I) which encode
CC prostate-specific proteins (II). (I) and (II) have cytostatic activity,
CC and can be used in vaccine production and gene therapy. (I), (II),
CC antibodies to (II), fusion proteins comprising (II), and isolated
CC T cells prepared using (I) or (II) are used treat cancer in a patient.
CC (I) and the antibodies are also used in the detection of cancer in a
CC patient. The cancer that is diagnosed or treated is particularly
CC prostate cancer. (I) and (II) can be used in vaccines. The antibodies or
CC (I) can be used for monitoring the progression of cancer in a patient.
CC (I) and (II) can also be used to improve diagnostic and therapeutic
CC methods for prostate cancer. They can indicate the level of metastasis
CC as well as the prostate volume. AAH93357 to AAH93944 and AAM01115 to
CC AAM01318 represent polynucleotide and amino acid sequences used in the
CC exemplification of the present invention.
XX
SQ Sequence 27 AA;

Query Match 46.4%; Score 13; DB 22; Length 27;
Best Local Similarity 100.0%; Pred. No. 6.4e-06;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 VGEGLYQGVPRAE 15
||||||||||||
Db 1 vgeglyqgvprae 13

RESULT 9
AAU69813
ID AAU69813 standard; Peptide; 15 AA.
XX
AC AAU69813;
XX
DT 30-JAN-2002 (first entry)
XX
DE Human prostate protein P703P antigenic epitope #10.
XX
KW Human; prostate cancer; ss; cytostatic; immunostimulant; tumour;
KW antigen; epitope.
XX
OS Homo sapiens.
XX
PN WO200173032-A2.
XX
PD 04-OCT-2001.
XX
PF 27-MAR-2001; 2001WO-US09919.
XX
PR 27-MAR-2000; 2000US-0536857.
PR 09-MAY-2000; 2000US-0568100.
PR 12-MAY-2000; 2000US-0570737.
PR 13-JUN-2000; 2000US-0593793.
PR 27-JUN-2000; 2000US-0605783.
PR 10-AUG-2000; 2000US-0636215.
PR 29-AUG-2000; 2000US-0651236.
PR 06-SEP-2000; 2000US-0657279.
PR 02-OCT-2000; 2000US-0679426.
PR 10-OCT-2000; 2000US-0685166.
XX
PA (CORI-) CORIXA CORP.
XX
PI Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Kalos MD;
PI Fanger GR, Retter MW, Stolk JA, Day CH, Vedvick TS, Carter D;
PI Li SX, Wang A, Skeiky YAW, Hepler WT, Henderson RA;
XX
DR WPI; 2001-639232/73.
XX
PT New human prostate-specific polypeptides and polynucleotides useful for
PT the diagnosis and treatment of cancer, especially prostate cancer -
XX
PS Example 19; Page 403; 579pp; English.
XX
CC The invention relates to isolated prostate-specific
CC polynucleotides, polypeptides, fusion proteins of the polypeptides,
CC antibodies raised against the polypeptides (or antigenic epitopes
CC derived from them) and antigen-presenting cells expressing the
CC polypeptides. The antibodies are useful for detecting the presence of
CC cancer, especially prostate cancer. The polypeptides, polynucleotides and
CC the antigen-presenting cells are useful for stimulating and/or expanding
CC T cells specific for a tumour protein, and for inhibiting the development
CC of cancer especially prostate cancer. Compositions comprising the
CC polynucleotide and/or polypeptide are useful for stimulating an immune
CC response, and for treating cancer. The oligonucleotide is useful for
CC detecting cancer. The present sequence is a prostate specific
CC peptide of the invention. The peptides either represent antigenic
CC epitopes or domains of prostate specific proteins.
XX
SQ Sequence 15 AA;

Query Match 39.3%; Score 11; DB 22; Length 15;
Best Local Similarity 100.0%; Pred. No. 0.00034;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 16 GTEARRHYDEG 26
||||||||||||
Db 5 gtearrhydeg 15

RESULT 10
AAM01168
ID AAM01168 standard; Peptide; 15 AA.
XX
AC AAM01168;
XX
DT 04-OCT-2001 (first entry)
XX
DE Human prostate-specific peptide used in epitope mapping studies #25.
XX
KW Human; prostate cancer; prostate-specific; diagnosis; vaccine;
KW cytostatic; gene therapy; metastasis.
XX
OS Homo sapiens.
XX
PN WO200151633-A2.
XX
PD 19-JUL-2001.
XX
PF 16-JAN-2001; 2001WO-US01574.
XX
PR 14-JAN-2000; 2000US-0483672.
XX
PA (CORI-) CORIXA CORP.
XX
PI Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Reed SG;
PI Kalos MD, Fanger GR, Day CH, Retter MW, Stolk JA, Skeiky YAW;

PI Wang A, Meagher MJ;
XX
DR WPI; 2001-425873/45.
XX
PT New polynucleotide encoding a prostate-specific protein, for
PT diagnosing, monitoring and treating prostate cancer in a patient and
PT for use in vaccines -
XX
PS Example 19; Page 402; 543pp; English.
XX
CC The present invention describes polynucleotide sequences (I) which encode
CC prostate-specific proteins (II). (I) and (II) have cytostatic activity,
CC and can be used in vaccine production and gene therapy. (I), (II),
CC antibodies to (II), fusion proteins comprising (II), and isolated
CC T cells prepared using (I) or (II) are used treat cancer in a patient.
CC (I) and the antibodies are also used in the detection of cancer in a
CC patient. The cancer that is diagnosed or treated is particularly
CC prostate cancer. (I) and (II) can be used in vaccines. The antibodies or
CC (I) can be used for monitoring the progression of cancer in a patient.
CC (I) and (II) can also be used to improve diagnostic and therapeutic
CC methods for prostate cancer. They can indicate the level of metastasis
CC as well as the prostate volume. AAH93357 to AAH93944 and AAM01115 to
CC AAM01318 represent polynucleotide and amino acid sequences used in the
CC exemplification of the present invention.
XX
SQ Sequence 15 AA;

Query Match 39.3%; Score 11; DB 22; Length 15;
Best Local Similarity 100.0%; Pred. No. 0.00034;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 16 GTEARRHYDEG 26
||||||||||
Db 5 gtearrhydeg 15

RESULT 11
AAG99053
ID AAG99053 standard; Peptide; 15 AA.
XX
AC AAG99053;
XX
DT 25-SEP-2001 (first entry)
XX
DE Human prostate-specific epitope mapping peptide #25.
XX
KW Human; prostate cancer; therapy; diagnosis; cat eye syndrome;
KW chromosome 22q11.2; prostate-specific protein; chromosome 1;
KW prostate specific antigen; PSA.
XX
OS Homo sapiens.
XX
PN WO200134802-A2.
XX
PD 17-MAY-2001.
XX
PF 09-NOV-2000; 2000WO-US30904.
XX
PR 12-NOV-1999; 99US-0439313.
PR 18-NOV-1999; 99US-0443686.
XX
PA (CORI-) CORIXA CORP.
XX
PI Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Reed SG;
PI Kalos MD, Retter MW, Stolk JA, Day CH, Skeiky YAW, Wang A;
XX
DR WPI; 2001-308785/32.
XX
PT Isolated polypeptide comprising at least an immunogenic portion of a
PT prostate-specific protein, useful in the diagnosis and therapy of
PT prostate cancer -
XX
PS Example 19; Page 302; 325pp; English.
XX
CC The present invention describes an isolated polypeptide (P1) comprising
CC at least an immunogenic portion of a prostate-specific protein, or its
CC variant. Also described are polynucleotides (N1) encoding (P1). (P1) and
CC (N1) have cytostatic activity and can be used in vaccine production.
CC The polypeptides, nucleic acids and antibodies from the present
CC invention are useful in the diagnosis and therapy of prostate cancer.
CC Prostate specific genes P704P, P712P, P774P, P775P and B305D are located
CC in a genomic region on chromosome 22q11.2 known as the Cat Eye Syndrome
CC region. Prostate specific antigen (PSA) P501S was located on
CC chromosome 1. AAH84671 to AAH85143 and AAG99000 to AAG99077 represent
CC polynucleotide and polypeptide sequences used in the exemplification
CC of the present invention.
XX
SQ Sequence 15 AA;

Query Match 39.3%; Score 11; DB 22; Length 15;
Best Local Similarity 100.0%; Pred. No. 0.00034;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 16 GTEARRHYDEG 26
||||||||||
Db 5 gtearrhydeg 15

RESULT 12
AAU69814
ID AAU69814 standard; Peptide; 17 AA.
XX
AC AAU69814;
XX
DT 30-JAN-2002 (first entry)
XX
DE Human prostate protein P703P antigenic epitope #11.
XX
KW Human; prostate cancer; ss; cytostatic; immunostimulant; tumour;
KW antigen; epitope.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN WO200173032-A2.
XX
PD 04-OCT-2001.
XX
PF 27-MAR-2001; 2001WO-US09919.
XX
PR 27-MAR-2000; 2000US-0536857.
PR 09-MAY-2000; 2000US-0568100.
PR 12-MAY-2000; 2000US-0570737.
PR 13-JUN-2000; 2000US-0593793.
PR 27-JUN-2000; 2000US-0605783.
PR 10-AUG-2000; 2000US-0636215.
PR 29-AUG-2000; 2000US-0651236.
PR 06-SEP-2000; 2000US-0657279.
PR 02-OCT-2000; 2000US-0679426.
PR 10-OCT-2000; 2000US-0685166.
XX
PA (CORI-) CORIXA CORP.
XX
PI Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Kalos MD;
PI Fanger GR, Retter MW, Stolk JA, Day CH, Vedvick TS, Carter D;
PI Li SX, Wang A, Skeiky YAW, Hepler WT, Henderson RA;
XX
DR WPI; 2001-639232/73.
XX
PT New human prostate-specific polypeptides and polynucleotides useful for
PT the diagnosis and treatment of cancer, especially prostate cancer -
XX
PS Claim 2; Page 403; 579pp; English.
XX

CC The invention relates to isolated prostate-specific
CC polynucleotides, polypeptides, fusion proteins of the polypeptides,
CC antibodies raised against the polypeptides (or antigenic epitopes
CC derived from them) and antigen-presenting cells expressing the
CC polypeptides. The antibodies are useful for detecting the presence of
CC cancer, especially prostate cancer. The polypeptides, polynucleotides and
CC the antigen-presenting cells are useful for stimulating and/or expanding
CC T cells specific for a tumour protein, and for inhibiting the development
CC of cancer especially prostate cancer. Compositions comprising the
CC polynucleotide and/or polypeptide are useful for stimulating an immune
CC response, and for treating cancer. The oligonucleotide is useful for
CC detecting cancer. The present sequence is a prostate specific
CC peptide of the invention. The peptides either represent antigenic
CC epitopes or domains of prostate specific proteins.
XX
SQ Sequence 17 AA;

Query Match 21.4%; Score 6; DB 22; Length 17;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 16 GTEARR 21
||||||
Db 5 gtearr 10

RESULT 13
AAM01169
ID AAM01169 standard; Peptide; 17 AA.
XX
AC AAM01169;
XX
DT 04-OCT-2001 (first entry)
XX
DE Human prostate-specific peptide used in epitope mapping studies #26.
XX
KW Human; prostate cancer; prostate-specific; diagnosis; vaccine;
KW cytostatic; gene therapy; metastasis.
XX
OS Homo sapiens.
XX
PN WO200151633-A2.
XX
PD 19-JUL-2001.
XX
PF 16-JAN-2001; 2001WO-US01574.
XX
PR 14-JAN-2000; 2000US-0483672.
XX
PA (CORI-) CORIXA CORP.
XX
PI Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Reed SG;
PI Kalos MD, Fanger GR, Day CH, Retter MW, Stolk JA, Skeiky YAW;
PI Wang A, Meagher MJ;
XX
DR WPI; 2001-425873/45.
XX
PT New polynucleotide encoding a prostate-specific protein, for
PT diagnosing, monitoring and treating prostate cancer in a patient and
PT for use in vaccines -
XX
PS Claim 2; Page 402-403; 543pp; English.
XX
CC The present invention describes polynucleotide sequences (I) which encode
CC prostate-specific proteins (II). (I) and (II) have cytostatic activity,
CC and can be used in vaccine production and gene therapy. (I), (II),
CC antibodies to (II), fusion proteins comprising (II), and isolated
CC T cells prepared using (I) or (II) are used treat cancer in a patient.
CC (I) and the antibodies are also used in the detection of cancer in a
CC patient. The cancer that is diagnosed or treated is particularly
CC prostate cancer. (I) and (II) can be used in vaccines. The antibodies or
CC (I) can be used for monitoring the progression of cancer in a patient.
CC (I) and (II) can also be used to improve diagnostic and therapeutic
CC methods for prostate cancer. They can indicate the level of metastasis
CC as well as the prostate volume. AAH93357 to AAH93944 and AAM01115 to
CC AAM01318 represent polynucleotide and amino acid sequences used in the
CC exemplification of the present invention.
XX
SQ Sequence 17 AA;

Query Match 21.4%; Score 6; DB 22; Length 17;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 16 GTEARR 21
||||||
Db 5 gtearr 10

RESULT 14
AAG99054
ID AAG99054 standard; Peptide; 17 AA.
XX
AC AAG99054;
XX
DT 25-SEP-2001 (first entry)
XX
DE Human prostate-specific epitope mapping peptide #26.
XX
KW Human; prostate cancer; therapy; diagnosis; cat eye syndrome;
KW chromosome 22q11.2; prostate-specific protein; chromosome 1;
KW prostate specific antigen; PSA.
XX
OS Homo sapiens.
XX
PN WO200134802-A2.
XX
PD 17-MAY-2001.
XX
PF 09-NOV-2000; 2000WO-US30904.
XX
PR 12-NOV-1999; 99US-0439313.
PR 18-NOV-1999; 99US-0443686.
XX
PA (CORI-) CORIXA CORP.
XX
PI Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Reed SG;
PI Kalos MD, Retter MW, Stolk JA, Day CH, Skeiky YAW, Wang A;
XX
DR WPI; 2001-308785/32.
XX
PT Isolated polypeptide comprising at least an immunogenic portion of a
PT prostate-specific protein, useful in the diagnosis and therapy of
PT prostate cancer -
XX
PS Claim 3; Page 302; 325pp; English.
XX
CC The present invention describes an isolated polypeptide (P1) comprising
CC at least an immunogenic portion of a prostate-specific protein, or its
CC variant. Also described are polynucleotides (N1) encoding (P1). (P1) and
CC (N1) have cytostatic activity and can be used in vaccine production.
CC The polypeptides, nucleic acids and antibodies from the present
CC invention are useful in the diagnosis and therapy of prostate cancer.
CC Prostate specific genes P704P, P712P, P774P, P775P and B305D are located
CC in a genomic region on chromosome 22q11.2 known as the Cat Eye Syndrome
CC region. Prostate specific antigen (PSA) P501S was located on
CC chromosome 1. AAH84671 to AAH85143 and AAG99000 to AAG99077 represent
CC polynucleotide and polypeptide sequences used in the exemplification
CC of the present invention.
XX
SQ Sequence 17 AA;

Query Match 21.4%; Score 6; DB 22; Length 17;

Best Local Similarity 100.0%; Pred. No. 26;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      16 GTEARR 21
           ||||||
Db       5 gtearr 10
```

RESULT 15
AAE01151
ID AAE01151 standard; Protein; 8 AA.
XX
AC AAE01151;
XX
DT 17-JUL-2001 (first entry)
XX
DE Peptide #1 of peppermint GPP synthase small subunit from clone pMp13.18.
XX
KW Peppermint; cytostatic; geranyl diphosphate synthase large subunit; GPP;
KW prenyltransferase; terpenoid family; dimethylallyl diphosphate; DMAPP;
KW isopentenyl diphosphate; IPP; acyclic precursor; monoterpene synthase;
KW limonene synthase; therapy; cancer; gene therapy; clone pMp13.18.
XX
OS Mentha piperita.
XX
PN WO200129188-A1.
XX
PD 26-APR-2001.
XX
PF 16-OCT-2000; 2000WO-US28690.
XX
PR 18-OCT-1999; 99US-0420211.
XX
PA (UNIW) UNIV WASHINGTON STATE RES FOUND.
XX
PI Croteau RB, Burke CC, Wildung MR;
XX
DR WPI; 2001-300331/31.
XX
PT New isolated recombinant geranyl diphosphate synthase large subunit
PT protein, useful for treating cancer -
XX
PS Example 1; Page 36; 68pp; English.
XX
CC The present sequence is a peptide of peppermint geranyl diphosphate
CC (GPP) synthase small subunit from clone pMp13.18. GPP synthase is one
CC of a family of enzymes called prenyltransferases that catalyse C5
CC elongation reactions to form the linear (acyclic) precursors of the
CC various terpenoid families. GPP synthase catalyses the condensation of
CC dimethylallyl diphosphate (DMAPP) and isopentenyl diphosphate (IPP) to
CC form GPP which is the intermediate, C10 acyclic precursor of the
CC monoterpenes. GPP synthase along with a monoterpene synthase protein,
CC preferably limonene synthase, which is capable of converting geranyl
CC diphosphate to a monoterpene having anti-cancer properties, is useful for
CC treating cancer in a mammalian host by introducing GPP synthase and the
CC monoterpene synthase protein into a cancerous cell. GPP synthase large
CC subunit DNA along with DNAs encoding GPP synthase small subunit and
CC monoterpene synthase protein is useful for treating cancer. GPP synthase
CC DNA is used in gene therapy.
XX
SQ Sequence 8 AA;

Query Match 17.9%; Score 5; DB 22; Length 8;
Best Local Similarity 100.0%; Pred. No. 6.4e+05;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       6 GLYQG 10
           |||||
Db       2 glyqg 6
```

Search completed: June 26, 2002, 13:51:47
Job time: 165 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: June 26, 2002, 13:49:32 ; Search time 14.89 Seconds
(without alignments)
45.931 Million cell updates/sec

Title: US-09-838-785-25
Perfect score: 28
Sequence: 1 DFVGEGLYQGVPRAEGTEARRHYDEGVR 28

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 231628 seqs, 24425594 residues

Word size : 0

Total number of hits satisfying chosen parameters: 140943

Minimum DB seq length: 0
Maximum DB seq length: 28

Post-processing: Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	13	46.4	25	4	US-09-439-313-520	Sequence 520, App
2	13	46.4	27	4	US-09-071-710-39	Sequence 39, Appl
3	13	46.4	27	4	US-09-525-397-39	Sequence 39, Appl
4	13	46.4	27	4	US-09-439-313-566	Sequence 566, App
5	11	39.3	15	4	US-09-439-313-518	Sequence 518, App
6	6	21.4	17	4	US-09-439-313-519	Sequence 519, App
7	5	17.9	8	4	US-09-420-211-12	Sequence 12, Appl
8	5	17.9	9	2	US-08-951-924A-3	Sequence 3, Appli
9	5	17.9	11	3	US-08-592-500-42	Sequence 42, Appl
10	5	17.9	11	3	US-08-195-006-42	Sequence 42, Appl
11	5	17.9	11	5	PCT-US94-07644A-42	Sequence 42, Appl
12	5	17.9	14	1	US-08-277-299-1	Sequence 1, Appli
13	5	17.9	14	2	US-08-880-671-1	Sequence 1, Appli
14	5	17.9	14	4	US-08-853-910-5	Sequence 5, Appli
15	5	17.9	15	1	US-07-969-336-1	Sequence 1, Appli
16	5	17.9	15	2	US-08-815-953-1	Sequence 1, Appli
17	5	17.9	17	2	US-08-472-659-4	Sequence 4, Appli
18	5	17.9	17	2	US-08-474-661-4	Sequence 4, Appli
19	5	17.9	17	2	US-08-611-977-4	Sequence 4, Appli
20	5	17.9	18	1	US-07-920-281C-27	Sequence 27, Appl
21	5	17.9	18	4	US-08-466-277-27	Sequence 27, Appl
22	5	17.9	19	4	US-08-975-040-8	Sequence 8, Appli
23	5	17.9	26	2	US-08-563-892A-1	Sequence 1, Appli
24	5	17.9	26	2	US-08-563-892A-20	Sequence 20, Appl
25	5	17.9	26	2	US-08-563-892A-21	Sequence 21, Appl
26	5	17.9	26	2	US-08-563-892A-22	Sequence 22, Appl
27	4	14.3	4	2	US-08-248-839C-21	Sequence 21, Appl
28	4	14.3	4	2	US-08-623-833B-24	Sequence 24, Appl
29	4	14.3	5	6	5166318-10	Patent No. 5166318
30	4	14.3	6	1	US-08-211-747-4	Sequence 4, Appli
31	4	14.3	6	3	US-09-146-675-7	Sequence 7, Appli
32	4	14.3	8	1	US-08-178-570-54	Sequence 54, Appl
33	4	14.3	8	1	US-08-071-357-14	Sequence 14, Appl
34	4	14.3	8	1	US-08-634-060-40	Sequence 40, Appl
35	4	14.3	8	1	US-08-211-747-1	Sequence 1, Appli
36	4	14.3	8	2	US-08-807-030-29	Sequence 29, Appl
37	4	14.3	8	2	US-08-807-030-53	Sequence 53, Appl
38	4	14.3	8	3	US-08-369-643-54	Sequence 54, Appl
39	4	14.3	8	5	PCT-US95-00147-54	Sequence 54, Appl
40	4	14.3	9	1	US-08-467-083-1	Sequence 1, Appli
41	4	14.3	9	1	US-08-211-747-8	Sequence 8, Appli
42	4	14.3	9	1	US-08-414-417B-1	Sequence 1, Appli
43	4	14.3	9	2	US-08-486-348A-1	Sequence 1, Appli
44	4	14.3	9	2	US-08-468-545B-1	Sequence 1, Appli
45	4	14.3	9	3	US-08-159-339A-741	Sequence 741, App

ALIGNMENTS

```
RESULT 1
US-09-439-313-520
; Sequence 520, Application US/09439313
; Patent No. 6329505
; GENERAL INFORMATION:
;  APPLICANT: Xu, Jiangchun
;  APPLICANT:  Dillon, Davin C.
;  APPLICANT:  Mitcham, Jennifer L.
;  APPLICANT:  Harlocker, Susan Louise
;  APPLICANT:  Jiang Yuqui
;  APPLICANT:  Reed, Steven G.
;  APPLICANT:  Kalos, Michael
;  APPLICANT:  Fanger, Gary
;  APPLICANT:  Retter, Mark
;  APPLICANT:  Solk, John
;  APPLICANT:  Day, Craig
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
;  TITLE OF INVENTION:  DIAGNOSIS OF PROSTATE CANCER
;  FILE REFERENCE: 210121.427C9
;  CURRENT APPLICATION NUMBER: US/09/439,313
;  CURRENT FILING DATE:  1999-11-12
;  NUMBER OF SEQ ID NOS: 575
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 520
;   LENGTH: 25
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Made in a lab
US-09-439-313-520


  Query Match             46.4%;  Score 13;  DB 4;  Length 25;
  Best Local Similarity   100.0%;  Pred. No. 1.7e-06;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 VGEGLYQGVPRAE 15
           |||||||||||||
Db       1 VGEGLYQGVPRAE 13


RESULT 2
US-09-071-710-39
; Sequence 39, Application US/09071710
; Patent No. 6130043
;  GENERAL INFORMATION:
;    APPLICANT:  BILLING-MEDEL, PATRICIA
;    APPLICANT:  COHEN, MAURICE
;    APPLICANT:  COLPITTS, TRACEY L.
;    APPLICANT:  FRIEDMAN, PAULA N.
```

```
;    APPLICANT:  GORDON, JULIAN
;    APPLICANT:  GRANADOS, EDWARD N.
;    APPLICANT:  HODGES, STEVEN C.
;    APPLICANT:  KLASS, MICHAEL R.
;    APPLICANT:  KRATOCHVIL, JON D.
;    APPLICANT:  ROBERTS-RAPP, LISA
;    APPLICANT:  RUSSELL, JOHN C.
;    APPLICANT:  STROUPE, STEPHEN D.
;    TITLE OF INVENTION:  REAGENTS AND METHODS USEFUL
;    TITLE OF INVENTION:  FOR DETECTING DISEASES OF THE PROSTATE
;    NUMBER OF SEQUENCES:  41
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Abbott Laboratories
;      STREET:  100 Abbott Park Road
;      CITY:  Abbott Park
;      STATE:  IL
;      COUNTRY:  USA
;      ZIP:  60064-3500
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/071,710
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/850,713
;      FILING DATE:  02-MAY-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Becker, Cheryl L.
;      REGISTRATION NUMBER:  35,441
;      REFERENCE/DOCKET NUMBER:  6083.US.P1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  847/935-1729
;      TELEFAX:  847/938-2623
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  39:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  27 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  No. 6130043e
US-09-071-710-39


  Query Match             46.4%;  Score 13;  DB 4;  Length 27;
  Best Local Similarity   100.0%;  Pred. No. 1.8e-06;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      16 GTEARRHYDEGVR 28
           |||||||||||||
Db      12 GTEARRHYDEGVR 24


RESULT   3
US-09-525-397-39
; Sequence 39, Application US/09525397
; Patent No. 6252047
;  GENERAL INFORMATION:
;    APPLICANT:  BILLING-MEDEL, PATRICIA
;    APPLICANT:  COHEN, MAURICE
;    APPLICANT:  COLPITTS, TRACEY L.
;    APPLICANT:  FRIEDMAN, PAULA N.
;    APPLICANT:  GORDON, JULIAN
;    APPLICANT:  GRANADOS, EDWARD N.
;    APPLICANT:  HODGES, STEVEN C.
;    APPLICANT:  KLASS, MICHAEL R.
;    APPLICANT:  KRATOCHVIL, JON D.
;    APPLICANT:  ROBERTS-RAPP, LISA
;    APPLICANT:  RUSSELL, JOHN C.
;    APPLICANT:  STROUPE, STEPHEN D.
;    TITLE OF INVENTION:  REAGENTS AND METHODS USEFUL
;    TITLE OF INVENTION:  FOR DETECTING DISEASES OF THE PROSTATE
;    NUMBER OF SEQUENCES:  41
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Abbott Laboratories
;      STREET:  100 Abbott Park Road
;      CITY:  Abbott Park
;      STATE:  IL
;      COUNTRY:  USA
;      ZIP:  60064-3500
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/525,397
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  09/071,710
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Becker, Cheryl L.
;      REGISTRATION NUMBER:  35,441
;      REFERENCE/DOCKET NUMBER:  6083.US.P1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  847/935-1729
;      TELEFAX:  847/938-2623
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  39:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  27 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  No. 6252047e
US-09-525-397-39


  Query Match             46.4%;  Score 13;  DB 4;  Length 27;
  Best Local Similarity   100.0%;  Pred. No. 1.8e-06;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      16 GTEARRHYDEGVR 28
           |||||||||||||
Db      12 GTEARRHYDEGVR 24


RESULT   4
US-09-439-313-566
; Sequence 566, Application US/09439313
; Patent No. 6329505
; GENERAL INFORMATION:
;  APPLICANT: Xu, Jiangchun
;  APPLICANT:  Dillon, Davin C.
;  APPLICANT:  Mitcham, Jennifer L.
;  APPLICANT:  Harlocker, Susan Louise
;  APPLICANT:  Jiang Yuqui
;  APPLICANT:  Reed, Steven G.
;  APPLICANT:  Kalos, Michael
;  APPLICANT:  Fanger, Gary
;  APPLICANT:  Retter, Mark
;  APPLICANT:  Solk, John
;  APPLICANT:  Day, Craig
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
;  TITLE OF INVENTION:  DIAGNOSIS OF PROSTATE CANCER
;  FILE REFERENCE: 210121.427C9
;  CURRENT APPLICATION NUMBER: US/09/439,313
;  CURRENT FILING DATE:  1999-11-12
;  NUMBER OF SEQ ID NOS: 575
;  SOFTWARE: FastSEQ for Windows Version 3.0
```

```
; SEQ ID NO 566
;   LENGTH: 27
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-439-313-566


  Query Match             46.4%;  Score 13;  DB 4;  Length 27;
  Best Local Similarity   100.0%;  Pred. No. 1.8e-06;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      3 VGEGLYQGVPRAE 15
          |||||||||||||
Db      1 VGEGLYQGVPRAE 13


RESULT   5
US-09-439-313-518
; Sequence 518, Application US/09439313
; Patent No. 6329505
; GENERAL INFORMATION:
;  APPLICANT: Xu, Jiangchun
;  APPLICANT:  Dillon, Davin C.
;  APPLICANT:  Mitcham, Jennifer L.
;  APPLICANT:  Harlocker, Susan Louise
;  APPLICANT:  Jiang Yuqui
;  APPLICANT:  Reed, Steven G.
;  APPLICANT:  Kalos, Michael
;  APPLICANT:  Fanger, Gary
;  APPLICANT:  Retter, Mark
;  APPLICANT:  Solk, John
;  APPLICANT:  Day, Craig
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
;  TITLE OF INVENTION:  DIAGNOSIS OF PROSTATE CANCER
;  FILE REFERENCE: 210121.427C9
;  CURRENT APPLICATION NUMBER: US/09/439,313
;  CURRENT FILING DATE:  1999-11-12
;  NUMBER OF SEQ ID NOS: 575
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 518
;   LENGTH: 15
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Made in a lab
US-09-439-313-518


  Query Match             39.3%;  Score 11;  DB 4;  Length 15;
  Best Local Similarity   100.0%;  Pred. No. 9.4e-05;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     16 GTEARRHYDEG 26
          ||||||||||
Db      5 GTEARRHYDEG 15


RESULT   6
US-09-439-313-519
; Sequence 519, Application US/09439313
; Patent No. 6329505
; GENERAL INFORMATION:
;  APPLICANT: Xu, Jiangchun
;  APPLICANT:  Dillon, Davin C.
;  APPLICANT:  Mitcham, Jennifer L.
;  APPLICANT:  Harlocker, Susan Louise
;  APPLICANT:  Jiang Yuqui
;  APPLICANT:  Reed, Steven G.
;  APPLICANT:  Kalos, Michael
;  APPLICANT:  Fanger, Gary
;  APPLICANT:  Retter, Mark
;  APPLICANT:  Solk, John
;  APPLICANT:  Day, Craig
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
;  TITLE OF INVENTION:  DIAGNOSIS OF PROSTATE CANCER
;  FILE REFERENCE: 210121.427C9
;  CURRENT APPLICATION NUMBER: US/09/439,313
;  CURRENT FILING DATE:  1999-11-12
;  NUMBER OF SEQ ID NOS: 575
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 519
;   LENGTH: 17
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Made in a lab
US-09-439-313-519


  Query Match             21.4%;  Score 6;  DB 4;  Length 17;
  Best Local Similarity   100.0%;  Pred. No. 6.9;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     16 GTEARR 21
          ||||||
Db      5 GTEARR 10


RESULT   7
US-09-420-211-12
; Sequence 12, Application US/09420211B
; Patent No. 6303330
; GENERAL INFORMATION:
;  APPLICANT: Croteau, Rodney B
;  APPLICANT: Burke, Charles C
;  APPLICANT: Wildung, Mark R
;  TITLE OF INVENTION: GERANYL DIPHOSPHATE SYNTHASE LARGE SUBUNIT, AND METHODS
;  TITLE OF INVENTION: OF USE
;  FILE REFERENCE: WSUR14294
;  CURRENT APPLICATION NUMBER: US/09/420,211B
;  CURRENT FILING DATE: 1999-10-18
;  EARLIER APPLICATION NUMBER: PCT/US98/21772
;  EARLIER FILING DATE: 1998-10-15
;  EARLIER APPLICATION NUMBER: 08/951,924
;  EARLIER FILING DATE: 1997-10-16
;  NUMBER OF SEQ ID NOS: 15
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 12
;   LENGTH: 8
;   TYPE: PRT
;   ORGANISM: Mentha piperita
US-09-420-211-12


  Query Match             17.9%;  Score 5;  DB 4;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 1.7e+05;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      6 GLYQG 10
          |||||
Db      2 GLYQG 6


RESULT   8
US-08-951-924A-3
; Sequence 3, Application US/08951924A
; Patent No. 5876964
;  GENERAL INFORMATION:
;    APPLICANT:  Croteau, Rodney B
;    APPLICANT:  Wildung, Mark R
;    APPLICANT:  Burke, Charles
;    APPLICANT:  Gershenzon, Jonathan
;    TITLE OF INVENTION:  GERANYL DIPHOSPHATE SYNTHASE FROM MINT
;    TITLE OF INVENTION:  (MENTHA PIPERITA)
;    NUMBER OF SEQUENCES:  7
;    CORRESPONDENCE ADDRESS:
```

```
;       ADDRESSEE:  CHRISTENSEN, O'CONNOR, JOHNSON AND KINDNESS
;       STREET:  1420 FIFTH AVENUE
;       CITY:  SEATTLE
;       STATE:  WASHINGTON
;       COUNTRY:  USA
;       ZIP:  WA 98101-2347
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/951,924A
;       FILING DATE:
;       CLASSIFICATION:  514
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Shelton, Dennis K
;       REGISTRATION NUMBER:  26,997
;       REFERENCE/DOCKET NUMBER:  WSUR 1 11217
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  206 695 1718
;       TELEFAX:  206 224 0779
;   INFORMATION FOR SEQ ID NO:  3:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  9 amino acids
;       TYPE:  amino acid
;       STRANDEDNESS:  not relevant
;       TOPOLOGY:  not relevant
;     MOLECULE TYPE:  peptide
;     HYPOTHETICAL:  NO
;     ANTI-SENSE:  NO
;     FRAGMENT TYPE:  internal
;     ORIGINAL SOURCE:
;       ORGANISM:  Mentha spicata
US-08-951-924A-3


  Query Match             17.9%;  Score 5;  DB 2;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 1.7e+05;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6 GLYQG 10
           |||||
Db       2 GLYQG 6


RESULT   9
US-08-592-500-42
; Sequence 42, Application US/08592500
; Patent No. 6005089
;  GENERAL INFORMATION:
;     APPLICANT:  Lanza, Francois
;     APPLICANT:  Phillips, David R.
;     APPLICANT:  Cazenave, Jean-Pierre
;     TITLE OF INVENTION:  Platelet Glycoprotein V Gene and Uses
;     NUMBER OF SEQUENCES:  43
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Townsend and Townsend Khourie and Crew
;       STREET:  379 Lytton Avenue
;       CITY:  Palo Alto
;       STATE:  California
;       COUNTRY:  US
;       ZIP:  94301
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/592,500
;       FILING DATE:
;       CLASSIFICATION:  536
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/089,455
;       FILING DATE:  09-JUL-1993
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Dow, Karen B.
;       REGISTRATION NUMBER:  29,684
;       REFERENCE/DOCKET NUMBER:  12418-28
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (415) 326-2400
;       TELEFAX:  (415) 326-2422
;   INFORMATION FOR SEQ ID NO:  42:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  11 amino acids
;       TYPE:  amino acid
;       TOPOLOGY:  unknown
;     MOLECULE TYPE:  peptide
;     HYPOTHETICAL:  NO
;     FEATURE:
;       NAME/KEY:  Peptide
;       LOCATION:  1..11
;       OTHER INFORMATION:  /note= "Amino acid sequence of the
;       OTHER INFORMATION:  human plasma factor XIII (FXIII) thrombin cleavage
;       OTHER INFORMATION:  site."
;     FEATURE:
;       NAME/KEY:  Region
;       LOCATION:  6..8
;       OTHER INFORMATION:  /note= "Amino acid residues
;       OTHER INFORMATION:  identical to GPV."
US-08-592-500-42


  Query Match             17.9%;  Score 5;  DB 3;  Length 11;
  Best Local Similarity   100.0%;  Pred. No. 44;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       9 QGVPR 13
           |||||
Db       3 QGVPR 7


RESULT  10
US-08-195-006-42
; Sequence 42, Application US/08195006
; Patent No. 6083688
;  GENERAL INFORMATION:
;     APPLICANT:  Lanza, Francois
;     APPLICANT:  Phillips, David R.
;     APPLICANT:  Cazenave, Jean-Pierre
;     TITLE OF INVENTION:  Platelet Glycoprotein V Gene and Uses
;     NUMBER OF SEQUENCES:  43
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Townsend and Townsend Khourie and Crew
;       STREET:  379 Lytton Avenue
;       CITY:  Palo Alto
;       STATE:  California
;       COUNTRY:  US
;       ZIP:  94301
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/195,006
;       FILING DATE:  10-FEB-1994
;       CLASSIFICATION:  514
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US 08/089,455
;       FILING DATE:  09-JUL-1993
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Dow, Karen B.
;       REGISTRATION NUMBER:  29,684
;       REFERENCE/DOCKET NUMBER:  12418-28
;     TELECOMMUNICATION INFORMATION:
```

```
;       TELEPHONE:  (415) 326-2400
;       TELEFAX:  (415) 326-2422
;  INFORMATION FOR SEQ ID NO:  42:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  11 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  peptide
;    HYPOTHETICAL:  NO
;    FEATURE:
;      NAME/KEY:  Peptide
;      LOCATION:  1..11
;      OTHER INFORMATION:  /note= "Amino acid sequence of the
;      OTHER INFORMATION:  human plasma factor XIII (FXIII) thrombin cleavage
;      OTHER INFORMATION:  site."
;    FEATURE:
;      NAME/KEY:  Region
;      LOCATION:  6..8
;      OTHER INFORMATION:  /note= "Amino acid residues
;      OTHER INFORMATION:  identical to GPV."
US-08-195-006-42


  Query Match             17.9%;  Score 5;  DB 3;  Length 11;
  Best Local Similarity   100.0%;  Pred. No. 44;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       9 QGVPR 13
           |||||
Db       3 QGVPR 7


RESULT  11
PCT-US94-07644A-42
; Sequence 42, Application PC/TUS9407644A
;  GENERAL INFORMATION:
;    APPLICANT:  COR Therapeutics, Inc.
;    TITLE OF INVENTION:  Platelet Glycoprotein V Gene and Uses
;    NUMBER OF SEQUENCES:  43
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Townsend and Townsend Khourie and Crew
;      STREET:  379 Lytton Avenue
;      CITY:  Palo Alto
;      STATE:  California
;      COUNTRY:  US
;      ZIP:  94301
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US94/07644A
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Dow, Karen B.
;      REGISTRATION NUMBER:  29,684
;      REFERENCE/DOCKET NUMBER:  012418-003000
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 326-2400
;      TELEFAX:  (415) 326-2422
;  INFORMATION FOR SEQ ID NO:  42:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  11 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  peptide
;    HYPOTHETICAL:  NO
;    FEATURE:
;      NAME/KEY:  Peptide
;      LOCATION:  1..11
;      OTHER INFORMATION:  /note= "Amino acid sequence of the
;      OTHER INFORMATION:  human plasma factor XIII (FXIII) thrombin cleavage
;      OTHER INFORMATION:  site."
;    FEATURE:
;      NAME/KEY:  Region
;      LOCATION:  6..8
;      OTHER INFORMATION:  /note= "Amino acid residues
;      OTHER INFORMATION:  identical to GPV."
PCT-US94-07644A-42


  Query Match             17.9%;  Score 5;  DB 5;  Length 11;
  Best Local Similarity   100.0%;  Pred. No. 44;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       9 QGVPR 13
           |||||
Db       3 QGVPR 7


RESULT  12
US-08-277-299-1
; Sequence 1, Application US/08277299
; Patent No. 5658744
;  GENERAL INFORMATION:
;    APPLICANT:  Ochoa, Augusto C.
;    APPLICANT:  Young, Howard A.
;    APPLICANT:  Longo, Dan L.
;    APPLICANT:  Ghosh, Paritosh
;    APPLICANT:  Robb, Richard J.
;    APPLICANT:  Neville, Mary E.
;    TITLE OF INVENTION:  METHODS OF IDENTIFYING PATIENTS HAVING
;    TITLE OF INVENTION:  ALTERED IMMUNE STATUS
;    NUMBER OF SEQUENCES:  3
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  3000 K Street, N.W., Suite 500
;      CITY:  Washington, D.C.
;      COUNTRY:  USA
;      ZIP:  20007-5109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/277,299
;      FILING DATE:  22-JUL-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  SAXE, Bernhard D.
;      REGISTRATION NUMBER:  28,665
;      REFERENCE/DOCKET NUMBER:  40403/132/ONCO
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202)672-5300
;      TELEFAX:  (202)672-5399
;      TELEX:  904136
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  14 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-08-277-299-1


  Query Match             17.9%;  Score 5;  DB 1;  Length 14;
  Best Local Similarity   100.0%;  Pred. No. 54;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6 GLYQG 10
           |||||
Db       9 GLYQG 13
```

```
RESULT  13
US-08-880-671-1
; Sequence 1, Application US/08880671
; Patent No. 5965366
;  GENERAL INFORMATION:
;    APPLICANT:  Ochoa, Augusto C.
;    APPLICANT:  Young, Howard A.
;    APPLICANT:  Longo, Dan L.
;    APPLICANT:  Ghosh, Paritosh
;    APPLICANT:  Robb, Richard J.
;    APPLICANT:  Neville, Mary E.
;    TITLE OF INVENTION:  METHODS OF IDENTIFYING PATIENTS HAVING
;    TITLE OF INVENTION:  ALTERED IMMUNE STATUS
;    NUMBER OF SEQUENCES:  3
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Foley & Lardner
;      STREET:  3000 K Street, N.W., Suite 500
;      CITY:  Washington, D.C.
;      COUNTRY:  USA
;      ZIP:  20007-5109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/880,671
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/277,299
;      FILING DATE:  22-JUL-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  SAXE, Bernhard D.
;      REGISTRATION NUMBER:  28,665
;      REFERENCE/DOCKET NUMBER:  40403/132/ONCO
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202)672-5300
;      TELEFAX:  (202)672-5399
;      TELEX:  904136
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  14 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-08-880-671-1


  Query Match             17.9%;  Score 5;  DB 2;  Length 14;
  Best Local Similarity   100.0%;  Pred. No. 54;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      6 GLYQG 10
          |||||
Db      9 GLYQG 13


RESULT  14
US-08-853-910-5
; Sequence 5, Application US/08853910
; Patent No. 6291247
;  GENERAL INFORMATION:
;    APPLICANT:  Riopelle, Richard J.
;    APPLICANT:  Ross, Gregory M.
;    APPLICANT:  Dory, Magdalena I.
;    APPLICANT:  Weaver, Donald F.
;    APPLICANT:  Shamovsky, Igor L.
;    TITLE OF INVENTION:  Factors and Methods for Reducing Biological
;    TITLE OF INVENTION:  Activity of a Multimeric Protein and Methods of Screening
;    TITLE OF INVENTION:  Therefor
;    NUMBER OF SEQUENCES:  11
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Dowell & Dowell
;      STREET:  2001 Jefferson Davis Highway
;      CITY:  Arlington
;      STATE:  Virginia
;      COUNTRY:  United States of America
;      ZIP:  22202-3697
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30 (EPO)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/853,910
;      FILING DATE:  09-MAY-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  USSN 08/241462
;      FILING DATE:  11-MAY-1994
;      APPLICATION NUMBER:  USSN 60/010 328
;      FILING DATE:  09-NOV-1995
;      APPLICATION NUMBER:  USSN 08/745608
;      FILING DATE:  08-NOV-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  RALPH A. DOWELL
;      REGISTRATION NUMBER:  26868
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (703) 415-2555
;      TELEFAX:  (703) 415-2559
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  14 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
US-08-853-910-5


  Query Match             17.9%;  Score 5;  DB 4;  Length 14;
  Best Local Similarity   100.0%;  Pred. No. 54;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      11 VPRAE 15
           |||||
Db       5 VPRAE 9


RESULT  15
US-07-969-336-1
; Sequence 1, Application US/07969336
; Patent No. 5681729
;  GENERAL INFORMATION:
;    APPLICANT:  Kudo, Toshio
;    APPLICANT:  Hiyamuta, Shuichi
;    APPLICANT:  Tanedani, Toshiyuki
;    APPLICANT:  Kadota, Akihiko
;    TITLE OF INVENTION:  Methods for Producing Human Lymphocytes
;    TITLE OF INVENTION:  and Human Monoclonal Antibodies, and Human Monoclonal
;    TITLE OF INVENTION:  Antibodies Produced Thereby
;    NUMBER OF SEQUENCES:  1
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Birch, Stewart, Kolasch and Birch
;      STREET:  301 No. 5681729th Washington Street
;      CITY:  Falls Church
;      STATE:  Virginia
;      COUNTRY:  U.S.A.
;      ZIP:  22046-3487
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/969,336
```

; FILING DATE: 19921030
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Raymond C.
; REGISTRATION NUMBER: 21,066
; REFERENCE/DOCKET NUMBER: 760-159P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 241-1300
; TELEFAX: (703) 241-2848
; TELEX: 248345
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 15 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; FRAGMENT TYPE: N-terminal
US-07-969-336-1

Query Match 17.9%; Score 5; DB 1; Length 15;
Best Local Similarity 100.0%; Pred. No. 57;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       3 VGEGL 7
           |||||
Db      11 VGEGL 15
```

Search completed: June 26, 2002, 13:52:14
Job time: 162 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: June 26, 2002, 13:50:07 ; Search time 20.27 Seconds
(without alignments)
132.733 Million cell updates/sec

Title: US-09-838-785-25
Perfect score: 28
Sequence: 1 DFVGEGLYQGVPRAEGTEARRHYDEGVR 28

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 283138 seqs, 96089334 residues

Word size : 0

Total number of hits satisfying chosen parameters: 5685

Minimum DB seq length: 0
Maximum DB seq length: 28

Post-processing: Listing first 45 summaries

Database : PIR_71:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	4	14.3	8	2	B45800	serum albumin - do
2	4	14.3	11	2	D60691	phycobilisome 9K l
3	4	14.3	12	2	S47393	T-cell antigen rec
4	4	14.3	13	2	G22565	R-phycoerythrin ga
5	4	14.3	14	2	PH1677	Ig heavy chain V r
6	4	14.3	16	2	A36285	site-specific meth
7	4	14.3	16	2	PH0754	T-cell receptor be
8	4	14.3	17	2	A60317	glucagon-like pept
9	4	14.3	18	2	A24749	neuropeptide A - b
10	4	14.3	20	2	A60525	lysozyme (EC 3.2.1
11	4	14.3	20	2	A40451	dormancy-related p
12	4	14.3	20	2	I58192	glial fibrillary a
13	4	14.3	21	2	PH1688	Ig heavy chain V r
14	4	14.3	21	2	PH1690	Ig heavy chain V r
15	4	14.3	22	2	PH1678	Ig heavy chain V r
16	4	14.3	22	2	PH1679	Ig heavy chain V r
17	4	14.3	22	2	PH1680	Ig heavy chain V r
18	4	14.3	22	2	A37335	68K surface antige
19	4	14.3	22	2	S03518	T-cell receptor ga
20	4	14.3	22	2	PT0052	translation initia
21	4	14.3	23	2	PH1681	Ig heavy chain V r
22	4	14.3	23	2	PH1682	Ig heavy chain V r
23	4	14.3	23	2	PH1689	Ig heavy chain V r
24	4	14.3	23	2	PH1691	Ig heavy chain V r
25	4	14.3	23	2	PH1692	Ig heavy chain V r
26	4	14.3	23	2	PH1693	Ig heavy chain V r
27	4	14.3	23	2	PH1694	Ig heavy chain V r
28	4	14.3	24	2	PH1683	Ig heavy chain V r
29	4	14.3	24	2	PH1684	Ig heavy chain V r
30	4	14.3	24	2	PH1685	Ig heavy chain V r
31	4	14.3	25	2	PH1686	Ig heavy chain V r
32	4	14.3	25	2	PH1700	Ig heavy chain V r
33	4	14.3	25	2	PH1701	Ig heavy chain V r
34	4	14.3	26	2	PH1687	Ig heavy chain V r
35	4	14.3	26	2	PH1703	Ig heavy chain V r
36	4	14.3	26	2	PH1704	Ig heavy chain V r
37	4	14.3	26	2	S27117	macrophage migrati
38	4	14.3	27	4	JE0015	probable 2.8K prot
39	4	14.3	27	4	JS0104	probable 2.8K prot
40	3	10.7	5	2	B22565	R-phycoerythrin al
41	3	10.7	6	2	PT0512	T-cell receptor be
42	3	10.7	6	2	PT0648	T-cell receptor be
43	3	10.7	7	2	PH1408	Ig heavy chain V r
44	3	10.7	7	2	A11483	aspartate transami
45	3	10.7	7	2	PT0529	T-cell receptor be

ALIGNMENTS

RESULT 1
B45800
serum albumin - dog (fragment)
C;Species: Canis lupus familiaris (dog)
C;Date: 03-Jun-1993 #sequence_revision 03-Jun-1993 #text_change 31-Dec-1993
C;Accession: B45800
R;Carraway, R.E.; Cochrane, D.E.; Boucher, W.; Mitra, S.P.
J. Immunol. 143, 1680-1684, 1989
A;Title: Structures of histamine-releasing peptides formed by the action of acid prot
A;Reference number: A45800; MUID:89341406
A;Accession: B45800
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-8 <CAR>

Query Match 14.3%; Score 4; DB 2; Length 8;
Best Local Similarity 100.0%; Pred. No. 2.8e+05;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 19 ARRH 22
||||
Db 2 ARRH 5

RESULT 2
D60691
phycobilisome 9K linker protein - Synechococcus sp. (PCC 7002) (fragment)
C;Species: Synechococcus sp.
C;Date: 14-May-1993 #sequence_revision 14-May-1993 #text_change 07-May-1999
C;Accession: D60691
R;Bryant, D.A.; de Lorimier, R.; Guglielmi, G.; Stevens Jr., S.E.
Arch. Microbiol. 153, 550-560, 1990
A;Title: Structural and compositional analyses of the phycobilisomes of Synechococcus s.
A;Reference number: A60691; MUID:90314662
A;Accession: D60691
A;Molecule type: protein
A;Residues: 1-11 <BRY>
C;Comment: This protein, one of the eleven components detected in this species of the
C;Keywords: photosystem II

Query Match 14.3%; Score 4; DB 2; Length 11;
Best Local Similarity 100.0%; Pred. No. 4.9e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 16 GTEA 19
||||
Db 8 GTEA 11

RESULT 3
S47393
T-cell antigen receptor VJ junction beta chain - human
C;Species: Homo sapiens (man)
C;Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 05-Nov-1999
C;Accession: S47393
R;Lehner, P.J.
submitted to the EMBL Data Library, August 1994
A;Description: Human HLA-A0201 restricted recognition of influenza A is dominated by T c
A;Reference number: S47355
A;Accession: S47393
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-12 <LEH>
A;Cross-references: EMBL:Z35670; NID:g527445; PIDN:CAA84739.1; PID:g527446
C;Keywords: T-cell receptor

Query Match 14.3%; Score 4; DB 2; Length 12;
Best Local Similarity 100.0%; Pred. No. 5.3e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 16 GTEA 19
||||
Db 7 GTEA 10

RESULT 4
G22565
R-phycoerythrin gamma-B chain - red alga (Gastroclonium coulteri) (fragment)
C;Species: Gastroclonium coulteri
C;Date: 07-Mar-1988 #sequence_revision 07-Mar-1988 #text_change 23-Mar-1993
C;Accession: G22565
R;Klotz, A.V.; Glazer, A.N.
J. Biol. Chem. 260, 4856-4863, 1985
A;Title: Characterization of the bilin attachment sites in R-phycoerythrin.
A;Reference number: A22565; MUID:85182601
A;Accession: G22565
A;Molecule type: protein
A;Residues: 1-13 <KLO>

Query Match 14.3%; Score 4; DB 2; Length 13;
Best Local Similarity 100.0%; Pred. No. 5.7e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 14 AEGT 17
||||
Db 2 AEGT 5

RESULT 5
PH1677
Ig heavy chain V region (clone NP-6-6) - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 24-Feb-1994 #sequence_revision 24-Feb-1994 #text_change 17-Mar-1999
C;Accession: PH1677
R;McHeyzer-Williams, M.G.; McLean, M.J.; Lalor, P.A.; Nossal, G.J.V.
J. Exp. Med. 178, 295-307, 1993
A;Title: Antigen-driven B cell differentiation in vivo.
A;Reference number: PH1675; MUID:93301607
A;Accession: PH1677
A;Molecule type: mRNA
A;Residues: 1-14 <MCH>
A;Experimental source: B cell
C;Superfamily: immunoglobulin V region; immunoglobulin homology
C;Keywords: heterotetramer; immunoglobulin

Query Match 14.3%; Score 4; DB 2; Length 14;
Best Local Similarity 100.0%; Pred. No. 6e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 17 TEAR 20
||||
Db 11 TEAR 14

RESULT 6
A36285
site-specific methyltransferase (EC 2.1.1.-) - Escherichia coli (fragment)
C;Species: Escherichia coli
C;Date: 18-Jan-1991 #sequence_revision 18-Jan-1991 #text_change 21-Mar-1996
C;Accession: A36285
R;Reich, N.O.; Everett, E.A.
J. Biol. Chem. 265, 8929-8934, 1990
A;Title: Identification of peptides involved in S-adenosylmethionine binding in the E
A;Reference number: A36285; MUID:90256827
A;Accession: A36285
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-16 <REI>
C;Keywords: methyltransferase

Query Match 14.3%; Score 4; DB 2; Length 16;
Best Local Similarity 100.0%; Pred. No. 6.8e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 16 GTEA 19
||||
Db 1 GTEA 4

RESULT 7
PH0754
T-cell receptor beta chain (F1) - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 17-Jul-1992 #sequence_revision 17-Jul-1992 #text_change 05-Nov-1999
C;Accession: PH0754
R;Casanova, J.L.; Romero, P.; Widmann, C.; Kourilsky, P.; Maryanski, J.L.
J. Exp. Med. 174, 1371-1383, 1991
A;Title: T cell receptor genes in a series of class I major histocompatibility comple
allelic exclusion and antigen-specific repertoire.
A;Reference number: PH0746; MUID:92078846
A;Accession: PH0754
A;Molecule type: mRNA
A;Residues: 1-16 <CAS>
A;Cross-references: EMBL:X60847; NID:g50935; PIDN:CAA43238.1; PID:g50936
A;Experimental source: T lymphocyte
C;Keywords: T-cell receptor

Query Match 14.3%; Score 4; DB 2; Length 16;
Best Local Similarity 100.0%; Pred. No. 6.8e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 24 DEGV 27
||||
Db 5 DEGV 8

RESULT 8
A60317
glucagon-like peptide 1 - marbled electric ray (fragment)
C;Species: Torpedo marmorata (marbled electric ray)
C;Date: 28-Oct-1992 #sequence_revision 28-Oct-1992 #text_change 21-Nov-1997
C;Accession: A60317
R;Conlon, J.M.; Hansen, H.F.; Schwartz, T.W.
Regul. Pept. 13, 94, 1986
A;Title: A truncated glucagon-like peptide I from torpedo pancreas.
A;Reference number: A60317
A;Accession: A60317
A;Molecule type: protein
A;Residues: 1-17 <CON>
C;Superfamily: glucagon

C;Keywords: duplication; pancreas

Query Match 14.3%; Score 4; DB 2; Length 17;
Best Local Similarity 100.0%; Pred. No. 7.1e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 14 AEGT 17
||||
Db 2 AEGT 5

RESULT 9
A24749
neuropeptide A - bovine
C;Species: Bos primigenius taurus (cattle)
C;Date: 28-Jul-1987 #sequence_revision 28-Jul-1987 #text_change 18-Aug-2000
C;Accession: A24749
R;Yang, H.Y.T.; Fratta, W.; Majane, E.A.; Costa, E.
Proc. Natl. Acad. Sci. U.S.A. 82, 7757-7761, 1985
A;Title: Isolation, sequencing, synthesis, and pharmacological characterization of two b
A;Reference number: A94074; MUID:86067985
A;Accession: A24749
A;Molecule type: protein
A;Residues: 1-18 <YAN>
C;Comment: The source of this peptide was brain.
C;Superfamily: unassigned animal peptides
C;Keywords: neuropeptide

Query Match 14.3%; Score 4; DB 2; Length 18;
Best Local Similarity 100.0%; Pred. No. 7.5e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 GEGL 7
||||
Db 2 GEGL 5

RESULT 10
A60525
lysozyme (EC 3.2.1.17) - cat (fragment)
C;Species: Felis silvestris catus (domestic cat)
C;Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 20-Jun-2000
C;Accession: A60525
R;Halliday, J.A.; Bell, K.; McKenzie, H.A.; Shaw, D.C.
Comp. Biochem. Physiol. B 95, 773-779, 1990
A;Title: Feline whey proteins: identification, isolation and initial characterization of
A;Reference number: A60525; MUID:90263403
A;Accession: A60525
A;Molecule type: protein
A;Residues: 1-20 <HAL>
A;Experimental source: milk
C;Superfamily: lysozyme c
C;Keywords: bacteriolytic enzyme; glycosidase; hydrolase; polysaccharide degradation

Query Match 14.3%; Score 4; DB 2; Length 20;
Best Local Similarity 100.0%; Pred. No. 8.2e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 13 RAEG 16
||||
Db 13 RAEG 16

RESULT 11
A40451
dormancy-related protein Pin 1 I - sugar pine (fragment)
C;Species: Pinus lambertiana (sugar pine)
C;Date: 06-Dec-1991 #sequence_revision 06-Dec-1991 #text_change 18-Jun-1993
C;Accession: A40451
R;Ekramoddoullah, A.K.M.
submitted to the Protein Sequence Database, November 1991
A;Reference number: A40451
A;Accession: A40451
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-20 <EKR>

Query Match 14.3%; Score 4; DB 2; Length 20;
Best Local Similarity 100.0%; Pred. No. 8.2e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 18 EARR 21
||||
Db 14 EARR 17

RESULT 12
I58192
glial fibrillary acidic protein GAFP - rat (fragment)
C;Species: Rattus sp. (rat)
C;Date: 27-Feb-1997 #sequence_revision 27-Feb-1997 #text_change 27-Feb-1997
C;Accession: I58192
R;Laping, N.J.; Morgan, T.E.; Nichols, N.R.; Rozovsky, I.; Young-Chan, C.S.; Zarow, C
Neuroscience 58, 563-572, 1994
A;Title: Transforming growth factor-beta 1 induces neuronal and astrocyte genes: tubu
A;Reference number: I58192; MUID:94224314
A;Accession: I58192
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-20 <RES>
A;Cross-references: GB:S70248; NID:g546523
A;Experimental source: strain Fisher
C;Genetics:
A;Introns: 11/2

Query Match 14.3%; Score 4; DB 2; Length 20;
Best Local Similarity 100.0%; Pred. No. 8.2e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 18 EARR 21
||||
Db 8 EARR 11

RESULT 13
PH1688
Ig heavy chain V region (clone NP-7-1) - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 24-Feb-1994 #sequence_revision 24-Feb-1994 #text_change 17-Mar-1999
C;Accession: PH1688
R;McHeyzer-Williams, M.G.; McLean, M.J.; Lalor, P.A.; Nossal, G.J.V.
J. Exp. Med. 178, 295-307, 1993
A;Title: Antigen-driven B cell differentiation in vivo.
A;Reference number: PH1675; MUID:93301607
A;Accession: PH1688
A;Molecule type: mRNA
A;Residues: 1-21 <MCH>
A;Experimental source: B cell
C;Superfamily: immunoglobulin V region; immunoglobulin homology
C;Keywords: heterotetramer; immunoglobulin

Query Match 14.3%; Score 4; DB 2; Length 21;
Best Local Similarity 100.0%; Pred. No. 8.5e+02;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 17 TEAR 20
||||
Db 11 TEAR 14

```
RESULT  14
PH1690
Ig heavy chain V region (clone NP-7-3) - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 24-Feb-1994 #sequence_revision 24-Feb-1994 #text_change 17-Mar-1999
C;Accession: PH1690
R;McHeyzer-Williams, M.G.; McLean, M.J.; Lalor, P.A.; Nossal, G.J.V.
J. Exp. Med. 178, 295-307, 1993
A;Title: Antigen-driven B cell differentiation in vivo.
A;Reference number: PH1675; MUID:93301607
A;Accession: PH1690
A;Molecule type: mRNA
A;Residues: 1-21 <MCH>
A;Experimental source: B cell
A;Note: the authors translated the codon AAT for residue 9 as Lys
C;Superfamily: immunoglobulin V region; immunoglobulin homology
C;Keywords: heterotetramer; immunoglobulin


  Query Match             14.3%;  Score 4;  DB 2;  Length 21;
  Best Local Similarity   100.0%;  Pred. No. 8.5e+02;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      17 TEAR 20
           ||||
Db      11 TEAR 14


RESULT  15
PH1678
Ig heavy chain V region (clone NP-6-9) - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 24-Feb-1994 #sequence_revision 24-Feb-1994 #text_change 17-Mar-1999
C;Accession: PH1678
R;McHeyzer-Williams, M.G.; McLean, M.J.; Lalor, P.A.; Nossal, G.J.V.
J. Exp. Med. 178, 295-307, 1993
A;Title: Antigen-driven B cell differentiation in vivo.
A;Reference number: PH1675; MUID:93301607
A;Accession: PH1678
A;Molecule type: mRNA
A;Residues: 1-22 <MCH>
A;Experimental source: B cell
C;Superfamily: immunoglobulin V region; immunoglobulin homology
C;Keywords: heterotetramer; immunoglobulin


  Query Match             14.3%;  Score 4;  DB 2;  Length 22;
  Best Local Similarity   100.0%;  Pred. No. 8.9e+02;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      17 TEAR 20
           ||||
Db      11 TEAR 14


Search completed: June 26, 2002, 13:52:47
Job time: 160 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: June 26, 2002, 13:51:08 ; Search time 9.93 Seconds
(without alignments)
109.179 Million cell updates/sec

Title: US-09-838-785-25
Perfect score: 28
Sequence: 1 DFVGEGLYQGVPRAEGTEARRHYDEGVR 28

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 105224 seqs, 38719550 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1673

Minimum DB seq length: 0
Maximum DB seq length: 28

Post-processing: Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	4	14.3	9	1	NEUX_HUMAN	P04277 homo sapien
2	4	14.3	18	1	NPA_BOVIN	P15506 bos taurus
3	4	14.3	20	1	FIBB_FELCA	P14469 felis silve
4	4	14.3	20	1	LYC_FELCA	P37155 felis silve
5	4	14.3	20	1	MIF_PIG	P80928 sus scrofa
6	4	14.3	21	1	FIBB_BISBO	P14466 bison bonas
7	4	14.3	21	1	FIBB_BUBBU	P14467 bubalus bub
8	4	14.3	21	1	FIBB_SYNCA	P14481 syncerus ca
9	4	14.3	21	1	OMP1_ACTPL	P80368 actinobacil
10	4	14.3	22	1	IF2G_PIG	P20461 sus scrofa
11	4	14.3	28	1	CH60_MYCSM	P80673 mycobacteri
12	3	10.7	5	1	TPIS_CANFA	P54714 canis famil
13	3	10.7	8	1	UPA1_HUMAN	P30087 homo sapien
14	3	10.7	8	1	VGLG_HSV2B	P81780 herpes simp
15	3	10.7	9	1	FIBB_PAPAN	P19344 papio anubi
16	3	10.7	9	1	FIBB_PAPHA	P19343 papio hamad
17	3	10.7	9	1	FIBB_THEGE	P19342 theropithec
18	3	10.7	9	1	LMT3_LOCMI	P41489 locusta mig
19	3	10.7	9	1	TKC1_CALVO	P41517 calliphora
20	3	10.7	9	1	TKL1_LOCMI	P16223 locusta mig
21	3	10.7	10	1	COXO_THUOB	P80982 thunnus obe
22	3	10.7	10	1	TKL2_LOCMI	P16224 locusta mig
23	3	10.7	10	1	TKL3_LOCMI	P30249 locusta mig
24	3	10.7	10	1	TKL4_LOCMI	P30250 locusta mig
25	3	10.7	10	1	TKNB_CHICK	P19851 gallus gall
26	3	10.7	10	1	TKNB_ONCMY	P28500 oncorhynchu
27	3	10.7	10	1	TKNK_PIG	P01292 sus scrofa
28	3	10.7	10	1	TKU1_UREUN	P40751 urechis uni
29	3	10.7	10	1	TPIS_NICPL	P19118 nicotiana p
30	3	10.7	10	1	TRP5_LEUMA	P81737 leucophaea
31	3	10.7	10	1	UP11_CAEEL	P55956 caenorhabdi
32	3	10.7	10	1	URA1_HUMAN	P32118 homo sapien
33	3	10.7	11	1	PKC1_CARMO	P82684 carausius m
34	3	10.7	11	1	TKC2_CALVO	P41518 calliphora
35	3	10.7	11	1	TKN1_PSEGU	P42986 pseudophryn
36	3	10.7	11	1	TKN2_PSEGU	P42987 pseudophryn
37	3	10.7	11	1	TKN3_PSEGU	P42988 pseudophryn
38	3	10.7	12	1	SOI5_BACSU	P80863 bacillus su
39	3	10.7	12	1	TKN1_KASMA	P08613 kassina mac
40	3	10.7	12	1	TKN_KASSE	P08611 kassina sen
41	3	10.7	13	1	FIBA_CAVPO	P14445 cavia porce
42	3	10.7	13	1	FIBB_HYLLA	P14472 hylobates l
43	3	10.7	13	1	NEUT_BUFMA	P81796 bufo marinu
44	3	10.7	13	1	NEUT_CHICK	P13724 gallus gall
45	3	10.7	13	1	NEUT_RANTE	P41536 rana tempor

ALIGNMENTS

RESULT 1
NEUX_HUMAN
ID NEUX_HUMAN STANDARD; PRT; 9 AA.
AC P04277;
DT 20-MAR-1987 (Rel. 04, Created)
DT 20-MAR-1987 (Rel. 04, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Neurotensin-related peptide (NRP) (Kinetensin).
OS Homo sapiens (Human),
OS Bos taurus (Bovine), and
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606, 9913, 9986;
RN [1]
RP SEQUENCE.
RC SPECIES=Human;
RX MEDLINE=86242180; PubMed=3087352;
RA Mogard M.H., Kobayashi R., Chen C.F., Lee T.D., Reeve J.R. Jr.,
RA Shively J.E., Walsh J.H.;
RT "The amino acid sequence of kinetensin, a novel peptide isolated from
RT pepsin-treated human plasma: homology with human serum albumin,
RT neurotensin and angiotensin.";
RL Biochem. Biophys. Res. Commun. 136:983-988(1986).
RN [2]
RP SEQUENCE.
RC SPECIES=Human, Bovine, and Rabbit;
RX MEDLINE=87194805; PubMed=2437111;
RA Carraway R.E., Mitra S.P., Cochrane D.E.;
RT "Structure of a biologically active neurotensin-related peptide
RT obtained from pepsin-treated albumin(s).";
RL J. Biol. Chem. 262:5968-5973(1987).
CC -!- FUNCTION: REGULATION OF FAT DIGESTION, LIPID ABSORPTION, AND
CC BLOOD FLOW (POTENTIAL).
DR PIR; A03239; ABHUSK.
DR PIR; A26693; A26693.
KW Hormone.
SQ SEQUENCE 9 AA; 1172 MW; C804DB4761F4140D CRC64;

Query Match 14.3%; Score 4; DB 1; Length 9;
Best Local Similarity 100.0%; Pred. No. 1e+05;
Matches 4; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 19 ARRH 22
||||
Db 2 ARRH 5

RESULT 2
NPA_BOVIN
ID NPA_BOVIN STANDARD; PRT; 18 AA.
AC P15506;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 01-APR-1990 (Rel. 14, Last annotation update)

```
DE   Neuropeptide A.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE.
RC   TISSUE=Brain;
RX   MEDLINE=86067985; PubMed=3865193;
RA   Yang H.-Y.T., Fratta W., Majane E.A., Costa E.;
RT   "Isolation, sequencing, synthesis, and pharmacological
RT   characterization of two brain neuropeptides that modulate the action
RT   of morphine.";
RL   Proc. Natl. Acad. Sci. U.S.A. 82:7757-7761(1985).
CC   -!- FUNCTION: MODULATES THE ACTION OF MORPHINE.
DR   PIR; A24749; A24749.
KW   Neuropeptide; Amidation.
FT   SIMILAR      15     18       TO NEUROPEPTIDE B (AA 5-8) (IDENTICAL).
FT   MOD_RES      18     18       AMIDATION.
SQ   SEQUENCE   18 AA;  1921 MW;  EC52DAE1F45CFCFB CRC64;


  Query Match             14.3%;  Score 4;  DB 1;  Length 18;
  Best Local Similarity   100.0%;  Pred. No. 2.7e+02;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       4 GEGL 7
           ||||
Db       2 GEGL 5


RESULT   3
FIBB_FELCA
ID   FIBB_FELCA     STANDARD;      PRT;     20 AA.
AC   P14469;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-JAN-1990 (Rel. 13, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Fibrinogen beta chain [Contains: Fibrinopeptide B] (Fragment).
GN   FGB.
OS   Felis silvestris catus (Cat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Carnivora; Fissipedia; Felidae; Felis.
OX   NCBI_TaxID=9685;
RN   [1]
RP   SEQUENCE.
RA   Blomback B., Blomback M., Grondahl N.J.;
RT   "Studies on fibrinopeptides from mammals.";
RL   Acta Chem. Scand. 19:1789-1791(1965).
CC   -!- FUNCTION: FIBRINOGEN HAS A DOUBLE FUNCTION: YIELDING MONOMERS THAT
CC       POLYMERIZE INTO FIBRIN AND ACTING AS A COFACTOR IN PLATELET
CC       AGGREGATION.
CC   -!- SUBUNIT: HEXAMER CONTAINING 2 SETS OF 3 NONIDENTICAL CHAINS
CC       (ALPHA, BETA AND GAMMA), LINKED TO EACH OTHER BY DISULFIDE BONDS.
CC   -!- MISCELLANEOUS: CONVERSION OF FIBRINOGEN TO FIBRIN IS TRIGGERED BY
CC       THROMBIN, WHICH CLEAVES FIBRINOPEPTIDES A AND B FROM ALPHA & BETA
CC       CHAINS, AND THUS EXPOSES THE N-TERMINAL POLYMERIZATION SITES
CC       RESPONSIBLE FOR THE FORMATION OF THE SOFT CLOT.
DR   InterPro; IPR002181; Fibrinogen_C.
DR   PROSITE; PS00514; FIBRIN_AG_C_DOMAIN; PARTIAL.
KW   Blood coagulation; Plasma.
FT   PEPTIDE       1     20       FIBRINOPEPTIDE B.
FT   NON_TER      20     20
SQ   SEQUENCE   20 AA;  2328 MW;  A829E393B8F627D0 CRC64;


  Query Match             14.3%;  Score 4;  DB 1;  Length 20;
  Best Local Similarity   100.0%;  Pred. No. 3e+02;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      23 YDEG 26
           ||||
Db       5 YDEG 8


RESULT   4
LYC_FELCA
ID   LYC_FELCA      STANDARD;      PRT;     20 AA.
AC   P37155;
DT   01-OCT-1994 (Rel. 30, Created)
DT   01-OCT-1994 (Rel. 30, Last sequence update)
DT   01-NOV-1997 (Rel. 35, Last annotation update)
DE   Lysozyme C (EC 3.2.1.17) (1,4-beta-N-acetylmuramidase C) (Fragment).
GN   LYZ.
OS   Felis silvestris catus (Cat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Carnivora; Fissipedia; Felidae; Felis.
OX   NCBI_TaxID=9685;
RN   [1]
RP   SEQUENCE.
RC   TISSUE=Milk;
RX   MEDLINE=90263403; PubMed=2344734;
RA   Halliday J.A., Bell K., McKenzie H.A., Shaw D.C.;
RT   "Feline whey proteins: identification, isolation and initial
RT   characterization of alpha-lactalbumin, beta-lactoglobulin and
RT   lysozyme.";
RL   Comp. Biochem. Physiol. 95B:773-779(1990).
CC   -!- FUNCTION: LYSOZYMES HAVE PRIMARILY BACTERIOLYTIC FUNCTION; THOSE
CC       IN TISSUES AND BODY FLUIDS ARE ASSOCIATED WITH THE MONOCYTE-
CC       MACROPHAGE SYSTEM AND ENHANCE THE ACTIVITY OF IMMUNOAGENTS.
CC   -!- CATALYTIC ACTIVITY: Hydrolysis of the 1,4-beta-linkages between N-
CC       acetyl-D-glucosamine and N-acetylmuramic acid in peptidoglycan
CC       heteropolymers of the prokaryotes cell walls.
CC   -!- SUBUNIT: MONOMER.
CC   -!- MISCELLANEOUS: LYSOZYME C IS CAPABLE OF BOTH HYDROLYSIS AND
CC       TRANSGLYCOSYLATION; IT SHOWS ALSO A SLIGHT ESTERASE ACTIVITY. IT
CC       ACTS RAPIDLY ON BOTH PEPTIDE-SUBSTITUTED AND UNSUBSTITUTED
CC       PEPTIDOGLYCAN &, SLOWLY, ON CHITIN OLIGOSACCHARIDES.
CC   -!- SIMILARITY: BELONGS TO FAMILY 22 OF GLYCOSYL HYDROLASES.
DR   PIR; A60525; A60525.
DR   HSSP; P11376; 2EQL.
DR   InterPro; IPR001916; Lactalbmn_lysozyme.
DR   Pfam; PF00062; lys; 1.
DR   PROSITE; PS00128; LACTALBUMIN_LYSOZYME; PARTIAL.
KW   Hydrolase; Glycosidase; Bacteriolytic enzyme; Milk.
FT   NON_TER      20     20
SQ   SEQUENCE   20 AA;  2314 MW;  EB8824EA59425E13 CRC64;


  Query Match             14.3%;  Score 4;  DB 1;  Length 20;
  Best Local Similarity   100.0%;  Pred. No. 3e+02;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      13 RAEG 16
           ||||
Db      13 RAEG 16


RESULT   5
MIF_PIG
ID   MIF_PIG        STANDARD;      PRT;     20 AA.
AC   P80928;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Macrophage migration inhibitory factor (MIF) (Phenylpyruvate
DE   tautomerase) (Glycosylation-inhibiting factor) (GIF) (Fragment).
GN   MIF.
OS   Sus scrofa (Pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX   NCBI_TaxID=9823;
RN   [1]
RP   SEQUENCE.
RA   Riviere S., Bouet F., Menez A., Galat A.;
```

```
RL   Submitted (MAR-1997) to the SWISS-PROT data bank.
CC   -!- FUNCTION: THE EXPRESSION OF MIF AT SITES OF INFLAMMATION SUGGEST A
CC       ROLE FOR THE MEDIATOR IN REGULATING THE FUNCTION OF MACROPHAGE IN
CC       HOST DEFENSE. ALSO ACTS AS A PHENYLPYRUVATE TAUTOMERASE (BY
CC       SIMILARITY).
CC   -!- SUBUNIT: HOMOTRIMER (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE MIF FAMILY.
DR   HSSP; P14174; 1GIF.
DR   InterPro; IPR001398; MIF.
DR   PROSITE; PS01158; MIF; PARTIAL.
KW   Isomerase; Macrophage; Inflammatory response; Cytokine.
FT   ACT_SITE      1      1       CATALYTIC BASE (BY SIMILARITY).
FT   NON_TER      20     20
SQ   SEQUENCE   20 AA;  2147 MW;  3517AF60F3012A61 CRC64;


  Query Match             14.3%;  Score 4;  DB 1;  Length 20;
  Best Local Similarity   100.0%;  Pred. No. 3e+02;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      11 VPRA 14
           ||||
Db       9 VPRA 12


RESULT   6
FIBB_BISBO
ID   FIBB_BISBO     STANDARD;      PRT;    21 AA.
AC   P14466;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-JAN-1990 (Rel. 13, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Fibrinogen beta chain [Contains: Fibrinopeptide B] (Fragment).
GN   FGB.
OS   Bison bonasus (European bison).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bison.
OX   NCBI_TaxID=9902;
RN   [1]
RP   SEQUENCE.
RA   Blomback B., Blomback M., Grondahl N.J.;
RT   "Studies on fibrinopeptides from mammals.";
RL   Acta Chem. Scand. 19:1789-1791(1965).
CC   -!- FUNCTION: FIBRINOGEN HAS A DOUBLE FUNCTION: YIELDING MONOMERS THAT
CC       POLYMERIZE INTO FIBRIN AND ACTING AS A COFACTOR IN PLATELET
CC       AGGREGATION.
CC   -!- SUBUNIT: HEXAMER CONTAINING 2 SETS OF 3 NONIDENTICAL CHAINS
CC       (ALPHA, BETA AND GAMMA), LINKED TO EACH OTHER BY DISULFIDE BONDS.
CC   -!- MISCELLANEOUS: CONVERSION OF FIBRINOGEN TO FIBRIN IS TRIGGERED BY
CC       THROMBIN, WHICH CLEAVES FIBRINOPEPTIDES A AND B FROM ALPHA & BETA
CC       CHAINS, AND THUS EXPOSES THE N-TERMINAL POLYMERIZATION SITES
CC       RESPONSIBLE FOR THE FORMATION OF THE SOFT CLOT.
DR   InterPro; IPR002181; Fibrinogen_C.
DR   PROSITE; PS00514; FIBRIN_AG_C_DOMAIN; PARTIAL.
KW   Blood coagulation; Plasma; Sulfation.
FT   PEPTIDE       1     21       FIBRINOPEPTIDE B.
FT   MOD_RES       6      6       SULFATION.
FT   NON_TER      21     21
SQ   SEQUENCE   21 AA;  2366 MW;  09EE75AF19E6363D CRC64;


  Query Match             14.3%;  Score 4;  DB 1;  Length 21;
  Best Local Similarity   100.0%;  Pred. No. 3.1e+02;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      23 YDEG 26
           ||||
Db       6 YDEG 9


RESULT   7
FIBB_BUBBU
ID   FIBB_BUBBU     STANDARD;      PRT;    21 AA.
AC   P14467;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Fibrinogen beta chain [Contains: Fibrinopeptide B] (Fragment).
GN   FGB.
OS   Bubalus bubalis (Domestic water buffalo).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bubalus.
OX   NCBI_TaxID=89462;
RN   [1]
RP   SEQUENCE.
RA   Mross G.A., Doolittle R.F.;
RT   "Amino acid sequence studies on artiodacty fibrinopeptides.";
RL   Arch. Biochem. Biophys. 122:674-684(1967).
RN   [2]
RP   SEQUENCE.
RC   STRAIN=ITALIAN BREED;
RX   MEDLINE=76040091; PubMed=1180969;
RA   Balestrieri C., Colonna G., Irace G.;
RT   "Covalent structure of fibrinopeptides from buffaloes breeding in
RT   Italy.";
RL   Biochim. Biophys. Acta 405:517-521(1975).
CC   -!- FUNCTION: FIBRINOGEN HAS A DOUBLE FUNCTION: YIELDING MONOMERS THAT
CC       POLYMERIZE INTO FIBRIN AND ACTING AS A COFACTOR IN PLATELET
CC       AGGREGATION.
CC   -!- SUBUNIT: HEXAMER CONTAINING 2 SETS OF 3 NONIDENTICAL CHAINS
CC       (ALPHA, BETA AND GAMMA), LINKED TO EACH OTHER BY DISULFIDE BONDS.
CC   -!- MISCELLANEOUS: CONVERSION OF FIBRINOGEN TO FIBRIN IS TRIGGERED BY
CC       THROMBIN, WHICH CLEAVES FIBRINOPEPTIDES A AND B FROM ALPHA & BETA
CC       CHAINS, AND THUS EXPOSES THE N-TERMINAL POLYMERIZATION SITES
CC       RESPONSIBLE FOR THE FORMATION OF THE SOFT CLOT.
DR   InterPro; IPR002181; Fibrinogen_C.
DR   PROSITE; PS00514; FIBRIN_AG_C_DOMAIN; PARTIAL.
KW   Blood coagulation; Plasma; Sulfation.
FT   PEPTIDE       1     21       FIBRINOPEPTIDE B.
FT   MOD_RES       1      1       PYRROLIDONE CARBOXYLIC ACID.
FT   MOD_RES       6      6       SULFATION.
FT   NON_TER      21     21
SQ   SEQUENCE   21 AA;  2379 MW;  09EE75BE4729163D CRC64;


  Query Match             14.3%;  Score 4;  DB 1;  Length 21;
  Best Local Similarity   100.0%;  Pred. No. 3.1e+02;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      23 YDEG 26
           ||||
Db       6 YDEG 9


RESULT   8
FIBB_SYNCA
ID   FIBB_SYNCA     STANDARD;      PRT;    21 AA.
AC   P14481;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Fibrinogen beta chain [Contains: Fibrinopeptide B] (Fragment).
GN   FGB.
OS   Syncerus caffer (Cape buffalo).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Syncerus.
OX   NCBI_TaxID=9970;
RN   [1]
RP   SEQUENCE.
RX   MEDLINE=67209145; PubMed=6033721;
RA   Doolittle R.F., Schubert D., Schwartz S.A.;
RT   "Amino acid sequence studies on artiodactyl fibrinopeptides. I.
RT   Dromedary camel, mule deer, and cape buffalo.";
```

```
RL   Arch. Biochem. Biophys. 118:456-467(1967).
CC   -!- FUNCTION: FIBRINOGEN HAS A DOUBLE FUNCTION: YIELDING MONOMERS THAT
CC       POLYMERIZE INTO FIBRIN AND ACTING AS A COFACTOR IN PLATELET
CC       AGGREGATION.
CC   -!- SUBUNIT: HEXAMER CONTAINING 2 SETS OF 3 NONIDENTICAL CHAINS
CC       (ALPHA, BETA AND GAMMA), LINKED TO EACH OTHER BY DISULFIDE BONDS.
CC   -!- MISCELLANEOUS: CONVERSION OF FIBRINOGEN TO FIBRIN IS TRIGGERED BY
CC       THROMBIN, WHICH CLEAVES FIBRINOPEPTIDES A AND B FROM ALPHA & BETA
CC       CHAINS, AND THUS EXPOSES THE N-TERMINAL POLYMERIZATION SITES
CC       RESPONSIBLE FOR THE FORMATION OF THE SOFT CLOT.
DR   InterPro; IPR002181; Fibrinogen_C.
DR   PROSITE; PS00514; FIBRIN_AG_C_DOMAIN; PARTIAL.
KW   Blood coagulation; Plasma; Sulfation.
FT   PEPTIDE       1     21       FIBRINOPEPTIDE B.
FT   MOD_RES       1      1       PYRROLIDONE CARBOXYLIC ACID.
FT   MOD_RES       6      6       SULFATION.
FT   NON_TER      21     21
SQ   SEQUENCE   21 AA;  2353 MW;  09EE75AF033B863D CRC64;


  Query Match             14.3%;  Score 4;  DB 1;  Length 21;
  Best Local Similarity   100.0%;  Pred. No. 3.1e+02;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      23 YDEG 26
           ||||
Db       6 YDEG 9


RESULT   9
OMP1_ACTPL
ID   OMP1_ACTPL     STANDARD;      PRT;     21 AA.
AC   P80368;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   40 kDa major outer membrane protein (MOMP) (Fragment).
OS   Actinobacillus pleuropneumoniae (Haemophilus pleuropneumoniae).
OC   Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC   Actinobacillus.
OX   NCBI_TaxID=715;
RN   [1]
RP   SEQUENCE.
RC   STRAIN=598;
RX   MEDLINE=96065432; PubMed=7483902;
RA   Hartmann L., Schroeder W., Luebke-Becker A.;
RT   "Isolation of the major outer-membrane protein of Actinobacillus
RT   pleuropneumoniae and Haemophilus parasuis.";
RL   J. Vet. Med. B 42:59-63(1995).
CC   -!- FUNCTION: STRUCTURAL RIGIDITY OF THE OUTER MEMBRANE OF ELEMENTARY
CC       BODIES AND PORIN FORMING, PERMITTING DIFFUSION OF SOLUTES THROUGH
CC       THE INTRACELLULAR RETICULATE BODY MEMBRANE.
CC   -!- SUBUNIT: DISULFIDE BOND INTERACTIONS WITHIN AND BETWEEN MOMP
CC       MOLECULES & OTHER COMPONENTS FORM HIGH MOLECULAR-WEIGHT OLIGOMERS.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane.
KW   Outer membrane; Transmembrane; Porin.
FT   NON_TER      21     21
SQ   SEQUENCE   21 AA;  2293 MW;  FFE7D12EA916563B CRC64;


  Query Match             14.3%;  Score 4;  DB 1;  Length 21;
  Best Local Similarity   100.0%;  Pred. No. 3.1e+02;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      14 AEGT 17
           ||||
Db       6 AEGT 9


RESULT  10
IF2G_PIG
ID   IF2G_PIG       STANDARD;      PRT;     22 AA.
AC   P20461;
DT   01-FEB-1991 (Rel. 17, Created)
DT   01-FEB-1991 (Rel. 17, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Eukaryotic translation initiation factor 2 subunit 3 (Eukaryotic
DE   translation initiation factor 2 gamma subunit) (eIF-2-gamma)
DE   (Fragment).
GN   EIF2S3 OR EIF2G.
OS   Sus scrofa (Pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX   NCBI_TaxID=9823;
RN   [1]
RP   SEQUENCE.
RC   TISSUE=Liver;
RA   Suzuki H., Mukouyama E.B.;
RT   "Pig liver translational initiation factor eIF-2: N-terminal amino
RT   acid sequences of alpha and gamma subunits and the phosphorylation
RT   site structure.";
RL   Agric. Biol. Chem. 52:1397-1408(1988).
CC   -!- FUNCTION: EIF-2 FUNCTIONS IN THE EARLY STEPS OF PROTEIN SYNTHESIS
CC       BY FORMING A TERNARY COMPLEX WITH GTP AND INITIATOR TRNA. THIS
CC       COMPLEX BINDS TO A 40S RIBOSOMAL SUBUNIT, FOLLOWED BY MRNA BINDING
CC       TO FORM A 43S PREINITIATION COMPLEX. JUNCTION OF THE 60S RIBOSOMAL
CC       SUBUNIT TO FORM THE 80S INITIATION COMPLEX IS PRECEDED BY
CC       HYDROLYSIS OF THE GTP BOUND TO EIF-2 AND RELEASE OF AN EIF-2-GDP
CC       BINARY COMPLEX. IN ORDER FOR EIF-2 TO RECYCLE AND CATALYZE ANOTHER
CC       ROUND OF INITIATION, THE GDP BOUND TO EIF-2 MUST EXCHANGE WITH GTP
CC       BY WAY OF A REACTION CATALYZED BY EIF-2B.
CC   -!- SUBUNIT: HETEROTRIMER COMPOSED OF AN ALPHA, A BETA, AND A GAMMA
CC       CHAIN.
CC   -!- SIMILARITY: BELONGS TO THE GTP-BINDING ELONGATION FACTOR FAMILY.
CC       EIF2G SUBFAMILY.
DR   PIR; PT0052; PT0052.
KW   Initiation factor; Protein biosynthesis; GTP-binding.
FT   NON_TER      22     22
SQ   SEQUENCE   22 AA;  2203 MW;  428B47D77D18B03C CRC64;


  Query Match             14.3%;  Score 4;  DB 1;  Length 22;
  Best Local Similarity   100.0%;  Pred. No. 3.2e+02;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      16 GTEA 19
           ||||
Db       2 GTEA 5


RESULT  11
CH60_MYCSM
ID   CH60_MYCSM     STANDARD;      PRT;     28 AA.
AC   P80673;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   60 kDa chaperonin (Protein Cpn60) (groEL protein) (Fragment).
GN   GROL OR MOPA OR GROEL.
OS   Mycobacterium smegmatis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1772;
RN   [1]
RP   SEQUENCE.
RC   STRAIN=ATCC 607 / MC(2)6 / NRRL B-692;
RX   MEDLINE=97387814; PubMed=9243799;
RA   Lundrigan M.D., Arceneaux J.E.L., Zhu W., Byers B.R.;
RT   "Enhanced hydrogen peroxide sensitivity and altered stress protein
RT   expression in iron-starved Mycobacterium smegmatis.";
RL   Biometals 10:215-225(1997).
CC   -!- FUNCTION: PREVENTS MISFOLDING AND PROMOTES THE REFOLDING AND
CC       PROPER ASSEMBLY OF UNFOLDED POLYPEPTIDES GENERATED UNDER STRESS
CC       CONDITIONS.
CC   -!- SUBUNIT: OLIGOMER OF 14 SUBUNITS COMPOSED OF TWO STACKED RINGS OF
CC       7 SUBUNITS (BY SIMILARITY).
```

```
CC   -!- SIMILARITY: BELONGS TO THE CHAPERONIN (HSP60) FAMILY.
DR   HSSP; P06139; 1AON.
DR   InterPro; IPR001844; Chaperonins_cpn60.
DR   PROSITE; PS00296; CHAPERONINS_CPN60; PARTIAL.
KW   Chaperone; ATP-binding.
FT   NON_TER      28     28
SQ   SEQUENCE   28 AA;  3047 MW;  2F40F27B94EF8720 CRC64;


  Query Match             14.3%;  Score 4;  DB 1;  Length 28;
  Best Local Similarity   100.0%;  Pred. No. 4e+02;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      18 EARR 21
           ||||
Db       9 EARR 12


RESULT  12
TPIS_CANFA
ID   TPIS_CANFA     STANDARD;      PRT;     5 AA.
AC   P54714;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   15-JUL-1998 (Rel. 36, Last annotation update)
DE   Triosephosphate isomerase (EC 5.3.1.1) (TIM) (Fragment).
GN   TPI1.
OS   Canis familiaris (Dog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX   NCBI_TaxID=9615;
RN   [1]
RP   SEQUENCE.
RC   TISSUE=Heart;
RX   MEDLINE=98163340; PubMed=9504812;
RA   Dunn M.J., Corbett J.M., Wheeler C.H.;
RT   "HSC-2DPAGE and the two-dimensional gel electrophoresis database of
RT   dog heart proteins.";
RL   Electrophoresis 18:2795-2802(1997).
CC   -!- CATALYTIC ACTIVITY: D-glyceraldehyde 3-phosphate = glycerone
CC       phosphate.
CC   -!- PATHWAY: PLAYS AN IMPORTANT ROLE IN SEVERAL METABOLIC PATHWAYS.
CC   -!- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO THE TRIOSEPHOSPHATE ISOMERASE FAMILY.
DR   HSC-2DPAGE; P54714; DOG.
DR   InterPro; IPR000652; Trioseph_isomerase.
DR   PROSITE; PS00171; TIM; PARTIAL.
KW   Isomerase; Glycolysis; Gluconeogenesis; Fatty acid biosynthesis;
KW   Pentose shunt.
FT   NON_TER       1      1
FT   NON_TER       5      5
SQ   SEQUENCE   5 AA;  550 MW;  64444862C9A00000 CRC64;


  Query Match             10.7%;  Score 3;  DB 1;  Length 5;
  Best Local Similarity   100.0%;  Pred. No. 1e+05;
  Matches    3;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 FVG 4
            |||
Db       1 FVG 3


RESULT  13
UPA1_HUMAN
ID   UPA1_HUMAN     STANDARD;      PRT;     8 AA.
AC   P30087;
DT   01-APR-1993 (Rel. 25, Created)
DT   01-APR-1993 (Rel. 25, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Unknown protein from 2D-page of plasma (Spot 2) (Fragment).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE.
RC   TISSUE=Plasma;
RX   MEDLINE=93092937; PubMed=1459097;
RA   Hughes G.J., Frutiger S., Paquet N., Ravier F., Pasquali C.,
RA   Sanchez J.-C., James R., Tissot J.-D., Bjellqvist B.,
RA   Hochstrasser D.F.;
RT   "Plasma protein map: an update by microsequencing.";
RL   Electrophoresis 13:707-714(1992).
CC   -!- MISCELLANEOUS: ON THE 2D-GEL THE DETERMINED PI OF THIS UNKNOWN
CC       PROTEIN IS: 4.9, ITS MW IS: 65 kDa.
DR   SWISS-2DPAGE; P30087; HUMAN.
FT   NON_TER       1      1
FT   UNSURE        8      8
FT   NON_TER       8      8
SQ   SEQUENCE   8 AA;  944 MW;  C01772C455BB06DA CRC64;


  Query Match             10.7%;  Score 3;  DB 1;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 1e+05;
  Matches    3;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      11 VPR 13
           |||
Db       6 VPR 8


RESULT  14
VGLG_HSV2B
ID   VGLG_HSV2B     STANDARD;      PRT;     8 AA.
AC   P81780;
DT   15-JUL-1999 (Rel. 38, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Glycoprotein G (Fragment).
OS   Herpes simplex virus (type 2 / strain B4327UR).
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Alphaherpesvirinae; Simplexvirus.
OX   NCBI_TaxID=103921;
RN   [1]
RP   SEQUENCE.
RA   Liljeqvist J.-A., Svennerholm B., Bergstrom T.;
RL   Submitted (APR-1999) to the SWISS-PROT data bank.
CC   -!- MISCELLANEOUS: THERE ARE SEVEN EXTERNAL GLYCOPROTEINS IN HSV1 AND
CC       2: GH, GB, GC, GG, GD, GI, AND GE.
CC   -!- MISCELLANEOUS: GLYCOPROTEIN G IS MUCH LARGER IN HSV-2 THAN IN
CC       HSV-1.
KW   Glycoprotein.
FT   NON_TER       8      8
SQ   SEQUENCE   8 AA;  683 MW;  7B47686772C865B8 CRC64;


  Query Match             10.7%;  Score 3;  DB 1;  Length 8;
  Best Local Similarity   100.0%;  Pred. No. 1e+05;
  Matches    3;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 GVP 12
           |||
Db       3 GVP 5


RESULT  15
FIBB_PAPAN
ID   FIBB_PAPAN     STANDARD;      PRT;     9 AA.
AC   P19344;
DT   01-NOV-1990 (Rel. 16, Created)
DT   01-NOV-1990 (Rel. 16, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Fibrinogen beta chain [Contains: Fibrinopeptide B] (Fragment).
GN   FGB.
OS   Papio anubis (Olive baboon).
```

```
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
OC   Cercopithecinae; Papio.
OX   NCBI_TaxID=9555;
RN   [1]
RP   SEQUENCE.
RX   MEDLINE=84161822; PubMed=6423621;
RA   Nakamura S., Takenaka O., Takahashi K.;
RT   "Fibrinopeptides A and B of baboons (Papio anubis, Papio hamadryas,
RT   and Theropithecus gelada): their amino acid sequences and
RT   evolutionary rates and a molecular phylogeny for the baboons.";
RL   J. Biochem. 94:1973-1978(1983).
CC   -!- FUNCTION: FIBRINOGEN HAS A DOUBLE FUNCTION: YIELDING MONOMERS THAT
CC       POLYMERIZE INTO FIBRIN AND ACTING AS A COFACTOR IN PLATELET
CC       AGGREGATION.
CC   -!- SUBUNIT: HEXAMER CONTAINING 2 SETS OF 3 NONIDENTICAL CHAINS
CC       (ALPHA, BETA AND GAMMA), LINKED TO EACH OTHER BY DISULFIDE BONDS.
CC   -!- MISCELLANEOUS: CONVERSION OF FIBRINOGEN TO FIBRIN IS TRIGGERED BY
CC       THROMBIN, WHICH CLEAVES FIBRINOPEPTIDES A AND B FROM ALPHA & BETA
CC       CHAINS, AND THUS EXPOSES THE N-TERMINAL POLYMERIZATION SITES
CC       RESPONSIBLE FOR THE FORMATION OF THE SOFT CLOT.
DR   PIR; D28854; D28854.
DR   InterPro; IPR002181; Fibrinogen_C.
DR   PROSITE; PS00514; FIBRIN_AG_C_DOMAIN; PARTIAL.
KW   Blood coagulation; Plasma.
FT   PEPTIDE       1      9       FIBRINOPEPTIDE B.
FT   NON_TER       9      9
SQ   SEQUENCE   9 AA;  1076 MW;  DDFE6409C7287B06 CRC64;


  Query Match             10.7%;  Score 3;  DB 1;  Length 9;
  Best Local Similarity   100.0%;  Pred. No. 1e+05;
  Matches    3;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5 EGL 7
           |||
Db       3 EGL 5


Search completed: June 26, 2002, 13:53:49
Job time: 161 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: June 26, 2002, 13:50:47 ; Search time 26.25 Seconds
(without alignments)
184.528 Million cell updates/sec

Title: US-09-838-785-25
Perfect score: 28
Sequence: 1 DFVGEGLYQGVPRAEGTEARRHYDEGVR 28

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 562222 seqs, 172994929 residues

Word size : 0

Total number of hits satisfying chosen parameters: 12285

Minimum DB seq length: 0
Maximum DB seq length: 28

Post-processing: Listing first 45 summaries

Database : SPTREMBL_19:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	5	17.9	23	4	Q9UC01	Q9uc01 homo sapien
2	5	17.9	26	12	Q9QRV0	Q9qrv0 hepatitis c
3	5	17.9	28	12	Q9QTJ2	Q9qtj2 tt virus. o
4	4	14.3	13	8	Q9T569	Q9t569 zea mays (m
5	4	14.3	14	2	Q9R8J2	Q9r8j2 chlamydia t
6	4	14.3	14	2	Q9R8J0	Q9r8j0 chlamydia t
7	4	14.3	14	2	Q9R8I8	Q9r8i8 chlamydia t
8	4	14.3	14	2	Q9R8I6	Q9r8i6 chlamydia t
9	4	14.3	14	2	Q9R8I4	Q9r8i4 chlamydia t
10	4	14.3	14	2	Q9R8I2	Q9r8i2 chlamydia t
11	4	14.3	14	2	Q9R8I0	Q9r8i0 chlamydia t
12	4	14.3	14	2	Q9R8H8	Q9r8h8 chlamydia t
13	4	14.3	14	2	Q9R8H6	Q9r8h6 chlamydia t
14	4	14.3	14	2	Q9R8H4	Q9r8h4 chlamydia t
15	4	14.3	14	2	Q9R8H2	Q9r8h2 chlamydia t
16	4	14.3	14	2	Q9R8H0	Q9r8h0 chlamydia t
17	4	14.3	14	2	Q9R8G8	Q9r8g8 chlamydia t
18	4	14.3	14	2	Q9R8G6	Q9r8g6 chlamydia t
19	4	14.3	14	2	Q9R8G4	Q9r8g4 chlamydia t
20	4	14.3	14	2	Q9R8G3	Q9r8g3 chlamydia t
21	4	14.3	14	2	Q9R8G1	Q9r8g1 chlamydia t
22	4	14.3	14	2	Q9R8F9	Q9r8f9 chlamydia t
23	4	14.3	14	2	Q9R8F7	Q9r8f7 chlamydia t
24	4	14.3	14	2	Q9R8F5	Q9r8f5 chlamydia t
25	4	14.3	14	2	Q9R8F2	Q9r8f2 chlamydia t
26	4	14.3	14	2	Q9R8F0	Q9r8f0 chlamydia t
27	4	14.3	14	2	Q9R8E9	Q9r8e9 chlamydia t
28	4	14.3	14	2	Q9R8E8	Q9r8e8 chlamydia t
29	4	14.3	14	2	Q9R8E7	Q9r8e7 chlamydia t
30	4	14.3	14	2	Q9R8E6	Q9r8e6 chlamydia t
31	4	14.3	14	2	Q9S3K8	Q9s3k8 chlamydia t
32	4	14.3	14	2	Q9S3K6	Q9s3k6 chlamydia t
33	4	14.3	16	2	Q9R4L1	Q9r4l1 spiroplasma
34	4	14.3	16	12	Q04246	Q04246 human parai
35	4	14.3	17	6	Q95M99	Q95m99 equus cabal
36	4	14.3	17	13	Q9PRU8	Q9pru8 gallus gall
37	4	14.3	19	2	O31045	O31045 streptomyce
38	4	14.3	19	8	Q95GI3	Q95gi3 abies alba
39	4	14.3	19	11	Q9QVN1	Q9qvn1 rattus sp.
40	4	14.3	20	4	Q9NZ26	Q9nz26 homo sapien
41	4	14.3	20	12	Q83272	Q83272 cucumber mo
42	4	14.3	20	13	Q9PRR4	Q9prr4 scyliorhinu
43	4	14.3	20	13	Q9PRR3	Q9prr3 scyliorhinu
44	4	14.3	22	10	Q9S899	Q9s899 pinus monti
45	4	14.3	22	11	Q9JHT2	Q9jht2 rattus norv

ALIGNMENTS

RESULT 1
Q9UC01
ID Q9UC01 PRELIMINARY; PRT; 23 AA.
AC Q9UC01;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE Y-23-R-NH2=PREDICTED INSULIN-LIKE GROWTH FACTOR IB-(103-124) E1 AMIDE
DE IBE1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92390398; PubMed=1325646;
RA Siegfried J.M., Kasprzyk P.G., Treston A.M., Mulshine J.L.,
RA Quinn K.A., Cuttitta F.;
RT "A mitogenic peptide amide encoded within the E peptide domain of the
RT insulin-like growth factor IB prohormone.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:8107-8111(1992).
SQ SEQUENCE 23 AA; 2508 MW; 40A5F3F520E3F789 CRC64;

Query Match 17.9%; Score 5; DB 4; Length 23;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      15 EGTEA 19
           |||||
Db      14 EGTEA 18
```

RESULT 2
Q9QRV0
ID Q9QRV0 PRELIMINARY; PRT; 26 AA.
AC Q9QRV0;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)

```
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   E2 GLYCOPROTEIN HYPERVARIABLE REGION (FRAGMENT).
OS   Hepatitis C virus.
OC   Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
OC   Hepacivirus.
OX   NCBI_TaxID=11103;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A-AS;
RA   Yeh C.-T.;
RT   "Replication of hepatitis C virus in the ascitic mononuclear cells and
RT   development of distinct quasispecies in the ascitic fluid.";
RL   Submitted (NOV-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF109739; AAD51570.1; -.
FT   NON_TER       1      1
FT   NON_TER      26     26
SQ   SEQUENCE   26 AA;  2462 MW;  945C5E047695C5CC CRC64;


  Query Match             17.9%;  Score 5;  DB 12;  Length 26;
  Best Local Similarity   100.0%;  Pred. No. 2.5e+02;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      16 GTEAR 20
           |||||
Db       7 GTEAR 11


RESULT   3
Q9QTJ2
ID   Q9QTJ2      PRELIMINARY;      PRT;    28 AA.
AC   Q9QTJ2;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE   ORF2 (FRAGMENT).
OS   TT virus.
OC   Viruses; ssDNA viruses; unclassified ssDNA viruses.
OX   NCBI_TaxID=68887;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C278;
RA   Gerner P.;
RT   "TT virus infection in healthy children and in children with chronic
RT   hepaititis B or C.";
RL   Submitted (SEP-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF188238; AAF04578.1; -.
FT   NON_TER      28     28
SQ   SEQUENCE   28 AA;  2733 MW;  ADEE944F7C15B22C CRC64;


  Query Match             17.9%;  Score 5;  DB 12;  Length 28;
  Best Local Similarity   100.0%;  Pred. No. 2.6e+02;
  Matches    5;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 GVPRA 14
           |||||
Db      19 GVPRA 23


RESULT   4
Q9T569
ID   Q9T569      PRELIMINARY;      PRT;    13 AA.
AC   Q9T569;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   MITOCHONDRIAL PLASMID S-2 DNA, 5' END (FRAGMENT).
OS   Zea mays (Maize).
OG   Mitochondrion.
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
OC   Panicoideae; Andropogoneae; Zea.
OX   NCBI_TaxID=4577;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=SEED;
RA   Traynor P.L., Levings C.S. III.;
RT   "Transcription of the S-2 maize mitochondrial plasmid.";
RL   Plant Mol. Biol. 7:255-263(1986).
DR   EMBL; M16992; AAA70281.1; -.
KW   Mitochondrion.
FT   NON_TER      13     13
SQ   SEQUENCE   13 AA;  1626 MW;  7B5E03673918A1F4 CRC64;


  Query Match             14.3%;  Score 4;  DB 8;  Length 13;
  Best Local Similarity   100.0%;  Pred. No. 1.4e+03;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      19 ARRH 22
           ||||
Db       4 ARRH 7


RESULT   5
Q9R8J2
ID   Q9R8J2      PRELIMINARY;      PRT;    14 AA.
AC   Q9R8J2;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   MAJOR OUTER MEMBRANE PROTEIN (FRAGMENT).
GN   OMP-1.
OS   Chlamydia trachomatis.
OC   Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX   NCBI_TaxID=813;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=REF A;
RA   Hsieh Y.-H., Bobo L.D.;
RT   "Diversity of major outer membrane protein (omp-1) of Chlamydia
RT   trachomatis in trachoma endemic villages, Kongwa, Tanzania.";
RL   Submitted (JUN-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF070309; AAC25280.1; -.
FT   NON_TER       1      1
FT   NON_TER      14     14
SQ   SEQUENCE   14 AA;  1405 MW;  14D7E247233CC4A0 CRC64;


  Query Match             14.3%;  Score 4;  DB 2;  Length 14;
  Best Local Similarity   100.0%;  Pred. No. 1.5e+03;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      16 GTEA 19
           ||||
Db      10 GTEA 13


RESULT   6
Q9R8J0
ID   Q9R8J0      PRELIMINARY;      PRT;    14 AA.
AC   Q9R8J0;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   MAJOR OUTER MEMBRANE PROTEIN (FRAGMENT).
GN   OMP-1.
OS   Chlamydia trachomatis.
OC   Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX   NCBI_TaxID=813;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A;
RA   Hsieh Y.-H., Bobo L.D.;
RT   "Diversity of major outer membrane protein (omp-1) of Chlamydia
```

```
RT   trachomatis in trachoma endemic villages, Kongwa, Tanzania.";
RL   Submitted (JUN-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF070313; AAC25284.1; -.
FT   NON_TER       1      1
FT   NON_TER      14     14
SQ   SEQUENCE   14 AA;  1405 MW;  14D7E247233CC4A0 CRC64;


  Query Match             14.3%;  Score 4;  DB 2;  Length 14;
  Best Local Similarity   100.0%;  Pred. No. 1.5e+03;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      16 GTEA 19
           ||||
Db      10 GTEA 13


RESULT   7
Q9R8I8
ID   Q9R8I8      PRELIMINARY;      PRT;    14 AA.
AC   Q9R8I8;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   MAJOR OUTER MEMBRANE PROTEIN (FRAGMENT).
GN   OMP-1.
OS   Chlamydia trachomatis.
OC   Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX   NCBI_TaxID=813;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=REF A12A;
RA   Hsieh Y.-H., Bobo L.D.;
RT   "Diversity of major outer membrane protein (omp-1) of Chlamydia
RT   trachomatis in trachoma endemic villages, Kongwa, Tanzania.";
RL   Submitted (JUN-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF070317; AAC25288.1; -.
FT   NON_TER       1      1
FT   NON_TER      14     14
SQ   SEQUENCE   14 AA;  1405 MW;  14D7E247233CC4A0 CRC64;


  Query Match             14.3%;  Score 4;  DB 2;  Length 14;
  Best Local Similarity   100.0%;  Pred. No. 1.5e+03;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      16 GTEA 19
           ||||
Db      10 GTEA 13


RESULT   8
Q9R8I6
ID   Q9R8I6      PRELIMINARY;      PRT;    14 AA.
AC   Q9R8I6;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   MAJOR OUTER MEMBRANE PROTEIN (FRAGMENT).
GN   OMP-1.
OS   Chlamydia trachomatis.
OC   Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX   NCBI_TaxID=813;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=REF A1A;
RA   Hsieh Y.-H., Bobo L.D.;
RT   "Diversity of major outer membrane protein (omp-1) of Chlamydia
RT   trachomatis in trachoma endemic villages, Kongwa, Tanzania.";
RL   Submitted (JUN-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF070321; AAC25292.1; -.
FT   NON_TER       1      1
FT   NON_TER      14     14
SQ   SEQUENCE   14 AA;  1405 MW;  14D7E247233CC4A0 CRC64;


  Query Match             14.3%;  Score 4;  DB 2;  Length 14;
  Best Local Similarity   100.0%;  Pred. No. 1.5e+03;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      16 GTEA 19
           ||||
Db      10 GTEA 13


RESULT   9
Q9R8I4
ID   Q9R8I4      PRELIMINARY;      PRT;    14 AA.
AC   Q9R8I4;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   MAJOR OUTER MEMBRANE PROTEIN (FRAGMENT).
GN   OMP-1.
OS   Chlamydia trachomatis.
OC   Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX   NCBI_TaxID=813;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=REF A14A;
RA   Hsieh Y.-H., Bobo L.D.;
RT   "Diversity of major outer membrane protein (omp-1) of Chlamydia
RT   trachomatis in trachoma endemic villages, Kongwa, Tanzania.";
RL   Submitted (JUN-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF070325; AAC25296.1; -.
FT   NON_TER       1      1
FT   NON_TER      14     14
SQ   SEQUENCE   14 AA;  1405 MW;  14D7E247233CC4A0 CRC64;


  Query Match             14.3%;  Score 4;  DB 2;  Length 14;
  Best Local Similarity   100.0%;  Pred. No. 1.5e+03;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      16 GTEA 19
           ||||
Db      10 GTEA 13


RESULT  10
Q9R8I2
ID   Q9R8I2      PRELIMINARY;      PRT;    14 AA.
AC   Q9R8I2;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   MAJOR OUTER MEMBRANE PROTEIN (FRAGMENT).
GN   OMP-1.
OS   Chlamydia trachomatis.
OC   Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX   NCBI_TaxID=813;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=REF A12B;
RA   Hsieh Y.-H., Bobo L.D.;
RT   "Diversity of major outer membrane protein (omp-1) of Chlamydia
RT   trachomatis in trachoma endemic villages, Kongwa, Tanzania.";
RL   Submitted (JUN-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF070329; AAC25300.1; -.
FT   NON_TER       1      1
FT   NON_TER      14     14
SQ   SEQUENCE   14 AA;  1405 MW;  14D7E247233CC4A0 CRC64;


  Query Match             14.3%;  Score 4;  DB 2;  Length 14;
  Best Local Similarity   100.0%;  Pred. No. 1.5e+03;
```

```
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      16 GTEA 19
           ||||
Db      10 GTEA 13


RESULT  11
Q9R8I0
ID   Q9R8I0      PRELIMINARY;      PRT;    14 AA.
AC   Q9R8I0;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   MAJOR OUTER MEMBRANE PROTEIN (FRAGMENT).
GN   OMP-1.
OS   Chlamydia trachomatis.
OC   Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX   NCBI_TaxID=813;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=REF A1B;
RA   Hsieh Y.-H., Bobo L.D.;
RT   "Diversity of major outer membrane protein (omp-1) of Chlamydia
RT   trachomatis in trachoma endemic villages, Kongwa, Tanzania.";
RL   Submitted (JUN-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF070333; AAC25304.1; -.
FT   NON_TER       1       1
FT   NON_TER      14      14
SQ   SEQUENCE   14 AA;  1405 MW;  14D7E247233CC4A0 CRC64;


  Query Match             14.3%;  Score 4;  DB 2;  Length 14;
  Best Local Similarity   100.0%;  Pred. No. 1.5e+03;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      16 GTEA 19
           ||||
Db      10 GTEA 13


RESULT  12
Q9R8H8
ID   Q9R8H8      PRELIMINARY;      PRT;    14 AA.
AC   Q9R8H8;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   MAJOR OUTER MEMBRANE PROTEIN (FRAGMENT).
GN   OMP-1.
OS   Chlamydia trachomatis.
OC   Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX   NCBI_TaxID=813;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=REF A1C;
RA   Hsieh Y.-H., Bobo L.D.;
RT   "Diversity of major outer membrane protein (omp-1) of Chlamydia
RT   trachomatis in trachoma endemic villages, Kongwa, Tanzania.";
RL   Submitted (JUN-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF070337; AAC25308.1; -.
FT   NON_TER       1       1
FT   NON_TER      14      14
SQ   SEQUENCE   14 AA;  1405 MW;  14D7E247233CC4A0 CRC64;


  Query Match             14.3%;  Score 4;  DB 2;  Length 14;
  Best Local Similarity   100.0%;  Pred. No. 1.5e+03;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      16 GTEA 19
           ||||
Db      10 GTEA 13


RESULT  13
Q9R8H6
ID   Q9R8H6      PRELIMINARY;      PRT;    14 AA.
AC   Q9R8H6;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   MAJOR OUTER MEMBRANE PROTEIN (FRAGMENT).
GN   OMP-1.
OS   Chlamydia trachomatis.
OC   Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX   NCBI_TaxID=813;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=REF A1D;
RA   Hsieh Y.-H., Bobo L.D.;
RT   "Diversity of major outer membrane protein (omp-1) of Chlamydia
RT   trachomatis in trachoma endemic villages, Kongwa, Tanzania.";
RL   Submitted (JUN-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF070341; AAC25312.1; -.
FT   NON_TER       1       1
FT   NON_TER      14      14
SQ   SEQUENCE   14 AA;  1405 MW;  14D7E247233CC4A0 CRC64;


  Query Match             14.3%;  Score 4;  DB 2;  Length 14;
  Best Local Similarity   100.0%;  Pred. No. 1.5e+03;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      16 GTEA 19
           ||||
Db      10 GTEA 13


RESULT  14
Q9R8H4
ID   Q9R8H4      PRELIMINARY;      PRT;    14 AA.
AC   Q9R8H4;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   MAJOR OUTER MEMBRANE PROTEIN (FRAGMENT).
GN   OMP-1.
OS   Chlamydia trachomatis.
OC   Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX   NCBI_TaxID=813;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=REF A1E;
RA   Hsieh Y.-H., Bobo L.D.;
RT   "Diversity of major outer membrane protein (omp-1) of Chlamydia
RT   trachomatis in trachoma endemic villages, Kongwa, Tanzania.";
RL   Submitted (JUN-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF070345; AAC25316.1; -.
FT   NON_TER       1       1
FT   NON_TER      14      14
SQ   SEQUENCE   14 AA;  1405 MW;  14D7E247233CC4A0 CRC64;


  Query Match             14.3%;  Score 4;  DB 2;  Length 14;
  Best Local Similarity   100.0%;  Pred. No. 1.5e+03;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      16 GTEA 19
           ||||
Db      10 GTEA 13


RESULT  15
Q9R8H2
ID   Q9R8H2      PRELIMINARY;      PRT;    14 AA.
```

```
AC   Q9R8H2;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   MAJOR OUTER MEMBRANE PROTEIN (FRAGMENT).
GN   OMP-1.
OS   Chlamydia trachomatis.
OC   Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX   NCBI_TaxID=813;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=REF A1F;
RA   Hsieh Y.-H., Bobo L.D.;
RT   "Diversity of major outer membrane protein (omp-1) of Chlamydia
RT   trachomatis in trachoma endemic villages, Kongwa, Tanzania.";
RL   Submitted (JUN-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF070349; AAC25320.1; -.
FT   NON_TER       1      1
FT   NON_TER      14     14
SQ   SEQUENCE   14 AA;  1405 MW;  14D7E247233CC4A0 CRC64;


  Query Match             14.3%;  Score 4;  DB 2;  Length 14;
  Best Local Similarity   100.0%;  Pred. No. 1.5e+03;
  Matches    4;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      16 GTEA 19
           ||||
Db      10 GTEA 13


Search completed: June 26, 2002, 13:53:26
Job time: 159 sec
```

PAGE BLANK

GenCore version 4.5

OM protein - protein search, using sw model

Run on: June 26, 2002, 13:43:07 ; Search time 31.66 Seconds
(without alignments)
98.233 Million cell updates/sec

Title: US-09-838-785-25
Perfect score: 28
Sequence: 1 DFVGEGLYQGVPRAEGTEARRHYDEGVR 28

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 747574 seqs, 111073796 residues

Word size : 0

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : A_Geneseq_032802:*
1: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	15	53.6	371	22	AAU69875	Human prostate cDN
2	15	53.6	371	22	AAM01230	P553S splice varia
3	15	53.6	371	22	AAE01362	Human gene 11 enco
4	15	53.6	553	19	AAW71869	Amino acid encoded
5	15	53.6	553	19	AAW69385	Prostate tumour sp
6	15	53.6	553	21	AAB28527	Protein encoded by
7	15	53.6	553	21	AAY82002	Human immunogenic
8	15	53.6	553	22	AAU69763	Human prostate cDN
9	15	53.6	553	22	AAU04961	Human prostate tum
10	15	53.6	553	22	AAM01117	Human prostate-spe
11	15	53.6	553	22	AAG99002	Human prostate-spe
12	15	53.6	553	22	AAG62150	Human P501S invent
13	15	53.6	553	22	AAB74800	Prostate tumour an
14	15	53.6	595	22	AAM01318	Alpha prepro-P501S
15	15	53.6	1079	22	AAB74830	Prostate tumour an
16	14	50.0	29	22	AAU69832	Human prostate pro
17	14	50.0	29	22	AAM01187	Human prostate-spe
18	14	50.0	29	22	AAG99072	Human prostate-spe
19	13	46.4	25	22	AAU69815	Human prostate pro
20	13	46.4	25	22	AAM01170	Human prostate-spe
21	13	46.4	25	22	AAG99055	Human prostate-spe
22	13	46.4	27	20	AAW85071	PS108 protein deri
23	13	46.4	27	21	AAB29271	Human prostate-rel
24	13	46.4	27	22	AAU69941	Human prostate pro
25	13	46.4	27	22	AAU04208	Prostate-specific
26	13	46.4	27	22	AAM01296	P501S peptide 296-
27	13	46.4	255	20	AAW85068	Protein encoded by
28	13	46.4	255	21	AAB29268	Human prostate-rel
29	13	46.4	255	22	AAU04205	Prostate-specific
30	11	39.3	15	22	AAU69813	Human prostate pro
31	11	39.3	15	22	AAM01168	Human prostate-spe
32	11	39.3	15	22	AAG99053	Human prostate-spe
33	7	25.0	179	22	ABB63079	Drosophila melanog
34	7	25.0	1958	15	AAR60620	Protein from ORF2
35	6	21.4	17	22	AAU69814	Human prostate pro
36	6	21.4	17	22	AAM01169	Human prostate-spe
37	6	21.4	17	22	AAG99054	Human prostate-spe
38	6	21.4	99	22	AAU43901	Propionibacterium
39	6	21.4	99	22	AAU65825	Propionibacterium
40	6	21.4	109	22	ABG00594	Novel human diagno
41	6	21.4	121	22	AAU66673	Propionibacterium
42	6	21.4	128	21	AAB43928	Human cancer assoc
43	6	21.4	150	22	ABB16915	Human nervous syst
44	6	21.4	150	22	AAM42375	Human polypeptide
45	6	21.4	157	22	AAB19929	Human oxidoreducta

ALIGNMENTS

RESULT 1
AAU69875
ID AAU69875 standard; Protein; 371 AA.
XX
AC AAU69875;
XX
DT 30-JAN-2002 (first entry)
XX
DE Human prostate cDNA encoded protein #69.
XX
KW Human; prostate cancer; cytostatic; immunostimulant; tumour; immunogen.
XX
OS Homo sapiens.
XX
PN WO200173032-A2.
XX
PD 04-OCT-2001.
XX
PF 27-MAR-2001; 2001WO-US09919.
XX
PR 27-MAR-2000; 2000US-0536857.
PR 09-MAY-2000; 2000US-0568100.
PR 12-MAY-2000; 2000US-0570737.
PR 13-JUN-2000; 2000US-0593793.
PR 27-JUN-2000; 2000US-0605783.
PR 10-AUG-2000; 2000US-0636215.
PR 29-AUG-2000; 2000US-0651236.
PR 06-SEP-2000; 2000US-0657279.
PR 02-OCT-2000; 2000US-0679426.
PR 10-OCT-2000; 2000US-0685166.
XX
PA (CORI-) CORIXA CORP.
XX
PI Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Kalos MD;

PI Fanger GR, Retter MW, Stolk JA, Day CH, Vedvick TS, Carter D;
PI Li SX, Wang A, Skeiky YAW, Hepler WT, Henderson RA;
XX
DR WPI; 2001-639232/73.
DR N-PSDB; AAS64038.
XX
PT New human prostate-specific polypeptides and polynucleotides useful for
PT the diagnosis and treatment of cancer, especially prostate cancer -
XX
PS Claim 2; Page 487-488; 579pp; English.
XX
CC The invention relates to isolated prostate-specific
CC polynucleotides, polypeptides, fusion proteins of the polypeptides,
CC antibodies raised against the polypeptides (or antigenic epitopes
CC derived from them) and antigen-presenting cells expressing the
CC polypeptides. The antibodies are useful for detecting the presence of
CC cancer, especially prostate cancer. The polypeptides, polynucleotides and
CC the antigen-presenting cells are useful for stimulating and/or expanding
CC T cells specific for a tumour protein, and for inhibiting the development
CC of cancer especially prostate cancer. Compositions comprising the
CC polynucleotide and/or polypeptide are useful for stimulating an immune
CC response, and for treating cancer. The oligonucleotide is useful for
CC detecting cancer. The present sequence is a prostate specific
CC polypeptide of the invention.
XX
SQ Sequence 371 AA;

Query Match 53.6%; Score 15; DB 22; Length 371;
Best Local Similarity 100.0%; Pred. No. 6.3e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DFVGEGLYQGVPRAE 15
||||||||||||||||
Db 268 dfvgeglyqgvprae 282

RESULT 2
AAM01230
ID AAM01230 standard; Protein; 371 AA.
XX
AC AAM01230;
XX
DT 04-OCT-2001 (first entry)
XX
DE P553S splice variant P553S-14 amino acid #2.
XX
KW Human; prostate cancer; prostate-specific; diagnosis; vaccine;
KW cytostatic; gene therapy; metastasis.
XX
OS Homo sapiens.
XX
PN WO200151633-A2.
XX
PD 19-JUL-2001.
XX
PF 16-JAN-2001; 2001WO-US01574.
XX
PR 14-JAN-2000; 2000US-0483672.
XX
PA (CORI-) CORIXA CORP.
XX
PI Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Reed SG;
PI Kalos MD, Fanger GR, Day CH, Retter MW, Stolk JA, Skeiky YAW;
PI Wang A, Meagher MJ;
XX
DR WPI; 2001-425873/45.
XX
PT New polynucleotide encoding a prostate-specific protein, for
PT diagnosing, monitoring and treating prostate cancer in a patient and
PT for use in vaccines -
XX
PS Claim 2; Page 464-466; 543pp; English.
XX
CC The present invention describes polynucleotide sequences (I) which encode
CC prostate-specific proteins (II). (I) and (II) have cytostatic activity,
CC and can be used in vaccine production and gene therapy. (I), (II),
CC antibodies to (II), fusion proteins comprising (II), and isolated
CC T cells prepared using (I) or (II) are used treat cancer in a patient.
CC (I) and the antibodies are also used in the detection of cancer in a
CC patient. The cancer that is diagnosed or treated is particularly
CC prostate cancer. (I) and (II) can be used in vaccines. The antibodies or
CC (I) can be used for monitoring the progression of cancer in a patient.
CC (I) and (II) can also be used to improve diagnostic and therapeutic
CC methods for prostate cancer. They can indicate the level of metastasis
CC as well as the prostate volume. AAH93357 to AAH93944 and AAM01115 to
CC AAM01318 represent polynucleotide and amino acid sequences used in the
CC exemplification of the present invention.
XX
SQ Sequence 371 AA;

Query Match 53.6%; Score 15; DB 22; Length 371;
Best Local Similarity 100.0%; Pred. No. 6.3e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DFVGEGLYQGVPRAE 15
||||||||||||||||
Db 268 dfvgeglyqgvprae 282

RESULT 3
AAE01362
ID AAE01362 standard; Protein; 371 AA.
XX
AC AAE01362;
XX
DT 18-JUL-2001 (first entry)
XX
DE Human gene 11 encoded secreted protein HWBAR14, SEQ ID NO:84.
XX
KW Human; secreted protein; proliferative disorder; cancer; tumour; asthma;
KW foetal abnormality; developmental abnormality; haematopoietic disorder;
KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
KW Parkinson's disease; cognitive disorder; schizophrenia; skin disorder;
KW psoriasis; sepsis; diabetes; atherosclerosis; cardiovascular disorder;
KW inflammation; neurological disorder; Alzheimer's disease; food additive;
KW angiogenic disorder; kidney disorder; gastrointestinal disorder; allergy;
KW pregnancy-related disorder; endocrine disorder; infection; wound healing;
KW cell culture; chemotaxis; vulnerary; binding partner identification;
KW gene therapy.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Peptide 1..48
FT /label= Signal_peptide
FT Protein 49..371
FT /label= Mature_human_secreted_protein
FT Misc-difference 20
FT /label= Unknown
FT /note= "Encoded by TNC"
XX
PN WO200134629-A1.
XX
PD 17-MAY-2001.
XX
PF 08-NOV-2000; 2000WO-US30654.
XX
PR 12-NOV-1999; 99US-0164835.
PR 27-JUL-2000; 2000US-0221142.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Ruben SM, Komatsoulis GA, Wei P, Fiscella M, Baker KP;
XX

DR WPI; 2001-308779/32.
DR N-PSDB; AAD05230.
XX
PT New nucleic acid encoding one of 21 human secreted proteins for
PT diagnosing, preventing, treating or ameliorating medical conditions,
PT such as autoimmune disease and cancer, and used as a food additive or
PT preservative -
XX
PS Claim 11; Page 438-440; 490pp; English.
XX
CC AAD05220-AAD05282 represent cDNAs corresponding to 21 human secreted
CC protein genes, and AAE01352-AAE01413 represent the proteins they encode.
CC AAE01415-AAE01433 represent human secreted protein fragments or variants.
CC The secreted proteins and their genes are useful for preventing, treating
CC or ameliorating medical conditions, e.g., by protein or gene therapy.
CC Pathological conditions can be diagnosed by determining the amount of the
CC new protein in a sample or by determining the presence of mutations in
CC the new genes. Specific uses are described for each of the 21 genes,
CC based on the tissues in which they are most highly expressed, and include
CC developing products for the diagnosis or treatment of proliferative
CC disorders, cancer, tumours, foetal and developmental abnormalities,
CC haematopoietic disorders, diseases of the immune system, AIDS, autoimmune
CC diseases (e.g., rheumatoid arthritis), inflammation, allergies,
CC neurological disorders (e.g., Alzheimer's disease, Parkinson's disease),
CC cognitive disorders, schizophrenia, asthma, skin disorders (e.g.,
CC psoriasis), sepsis, diabetes, atherosclerosis, cardiovascular disorders,
CC angiogenic disorders, kidney disorders, gastrointestinal disorders,
CC pregnancy-related disorders, endocrine disorders, and infections. The
CC proteins can also be used to aid wound healing and epithelial cell
CC proliferation, to prevent skin aging due to sunburn, to maintain organs
CC before transplantation, for supporting cell culture of primary tissues,
CC to regenerate tissues, to identify their cognate ligands or binding
CC partners, and in chemotaxis, and can be used as a food additive or
CC preservative to modify storage properties. Antibodies specific for a
CC protein of the invention can be used in alleviating symptoms associated
CC with the disorders mentioned above, and in diagnostic immunoassays e.g.,
CC radioimmunoassay or enzyme linked immunosorbent assay (ELISA).
CC The present sequence represents a human secreted protein of
CC the invention.
XX
SQ Sequence 371 AA;

Query Match 53.6%; Score 15; DB 22; Length 371;
Best Local Similarity 100.0%; Pred. No. 6.3e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DFVGEGLYQGVPRAE 15
||||||||||||||||
Db 268 dfvgeglyqgvprae 282

RESULT 4
AAW71869
ID AAW71869 standard; Protein; 553 AA.
XX
AC AAW71869;
XX
DT 06-JAN-1999 (first entry)
XX
DE Amino acid encoded by prostate tumour clone L1-12.
XX
KW Prostate; cancer; tumour; vaccine; immunogen; clone.
XX
OS Homo sapiens.
XX
PN WO9837093-A2.
XX
PD 27-AUG-1998.
XX
PF 25-FEB-1998; 98WO-US03492.
XX
PR 09-FEB-1998; 98US-0020956.
PR 25-FEB-1997; 97US-0806099.
PR 01-AUG-1997; 97US-0904804.
XX
PA (CORI-) CORIXA CORP.
XX
PI Dillon DC, Xu J;
XX
DR WPI; 1998-609886/51.
DR N-PSDB; AAV61201.
XX
PT Polypeptides comprising immunogenic portions of prostate proteins -
PT used in a vaccine for the treatment of prostate cancer
XX
PS Example 1; Page 82-84; 130pp; English.
XX
CC The present sequence is an immunogenic portion of a prostate tumour
CC protein. The immunogen, or the DNA encoding it, can be used as a
CC vaccine for the treatment of prostate cancer. The immunogen was
CC isolated from a prostate tumour cDNA library obtained by subtracting
CC a prostate tumour cDNA expression library with a normal tissue cDNA
CC library.
XX
SQ Sequence 553 AA;

Query Match 53.6%; Score 15; DB 19; Length 553;
Best Local Similarity 100.0%; Pred. No. 8.8e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DFVGEGLYQGVPRAE 15
||||||||||||||||
Db 294 dfvgeglyqgvprae 308

RESULT 5
AAW69385
ID AAW69385 standard; Protein; 553 AA.
XX
AC AAW69385;
XX
DT 08-DEC-1998 (first entry)
XX
DE Prostate tumour specific gene clone L1-12 protein.
XX
KW Prostate tumour specific gene; human; prostate cancer; detection;
KW therapy.
XX
OS Homo sapiens.
XX
PN WO9837418-A2.
XX
PD 27-AUG-1998.
XX
PF 25-FEB-1998; 98WO-US03690.
XX
PR 09-FEB-1998; 98US-0904809.
PR 25-FEB-1997; 97US-0806596.
PR 01-AUG-1997; 97US-0904809.
XX
PA (CORI-) CORIXA CORP.
XX
PI Dillon DC, Xu J;
XX
DR WPI; 1998-480805/41.
DR N-PSDB; AAV58586.
XX
PT Novel human prostate specific tumour protein and fragments - useful
PT for detecting and treating prostate cancers
XX
PS Example 1; Page 87-89; 141pp; English.
XX
CC This sequence is encoded by a human prostate tumour specific gene, and
CC can be used in the method of the invention. The method is for detecting

```
CC   prostate cancer comprises contacting a biological sample with an agent
CC   able to bind an immunogenic portion of a prostate protein (such as
CC   this protein sequence). An antibody which binds to an immunogenic
CC   portion of the prostate protein, and the method can be used to detect,
CC   monitor progression of, or treat prostate cancers. The antibody may
CC   also be conjugated to a therapeutic agent for use in therapy of prostate
CC   cancers.
XX
SQ   Sequence   553 AA;


  Query Match             53.6%;  Score 15;  DB 19;  Length 553;
  Best Local Similarity   100.0%;  Pred. No. 8.8e-07;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 DFVGEGLYQGVPRAE 15
           |||||||||||||||
Db     294 dfvgeglyqgvprae 308


RESULT   6
AAB28527
ID   AAB28527 standard; Protein; 553 AA.
XX
AC   AAB28527;
XX
DT   07-FEB-2001  (first entry)
XX
DE   Protein encoded by human breast tumour cDNA clone P501S.
XX
KW   Human; breast tumour antigen; cytostatic; immunotherapy;
KW   breast cancer; vaccine.
XX
OS   Homo sapiens.
XX
PN   WO200061756-A2.
XX
PD   19-OCT-2000.
XX
PF   10-APR-2000; 2000WO-US09688.
XX
PR   09-APR-1999;   99US-0288950.
PR   02-JUL-1999;   99US-0346327.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Reed SG,  Xu J,  Dillon DC;
XX
DR   WPI; 2000-638568/61.
DR   N-PSDB; AAC79473.
XX
PT   A novel isolated polypeptide comprising an immunogenic portion of a
PT   breast cancer protein useful in the detection and treatment of breast
PT   cancer -
XX
PS   Claim 2; Page 92-93; 95pp; English.
XX
CC   The present sequence is encoded by a cDNA sequence which was isolated
CC   from a breast tumour cDNA library. It is provided in a specification
CC   relating to compounds for immunotherapy and diagnosis of breast cancer.
CC   Breast tumour antigens and the polynucleotides that encode them may be
CC   used in the production of a pharmaceutical composition to be used in the
CC   treatment of breast cancer. Proliferated T cells and incubated antigen
CC   presenting cells are also required. The polypeptides and polynucleotides
CC   may also be used to produce a vaccine.
XX
SQ   Sequence   553 AA;


  Query Match             53.6%;  Score 15;  DB 21;  Length 553;
  Best Local Similarity   100.0%;  Pred. No. 8.8e-07;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 DFVGEGLYQGVPRAE 15
           |||||||||||||||
Db     294 dfvgeglyqgvprae 308


RESULT   7
AAY82002
ID   AAY82002 standard; Protein; 553 AA.
XX
AC   AAY82002;
XX
DT   13-JUN-2000  (first entry)
XX
DE   Human immunogenic prostate tumour protein sequence SEQ ID NO:113.
XX
KW   Human; prostate cancer; diagnosis; tumour; gene therapy; detection;
KW   immunogenic; cytostatic; vaccine.
XX
OS   Homo sapiens.
XX
PN   WO200004149-A2.
XX
PD   27-JAN-2000.
XX
PF   14-JUL-1999;   99WO-US15838.
XX
PR   14-JUL-1998;   98US-0115453.
PR   14-JUL-1998;   98US-0116134.
PR   23-SEP-1998;   98US-0159812.
PR   23-SEP-1998;   98US-0159822.
PR   15-JAN-1999;   99US-0232149.
PR   15-JAN-1999;   99US-0232880.
PR   09-APR-1999;   99US-0288946.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Dillon DC,  Harlocker SL,  Yuqiu J,  Xu J,  Mitcham JL;
XX
DR   WPI; 2000-171268/15.
XX
PT   New polypeptide useful for treating and diagnosing prostate cancer
PT   comprises an immunogenic portion of prostate tumor protein -
XX
PS   Claim 3; Page 138-139; 263pp; English.
XX
CC   The present invention describes isolated polypeptides, comprising an
CC   immunogenic portion of a prostate tumour protein (PTP). The polypeptides
CC   and polynucleotides encoding them have cytostatic activity and can be
CC   used in vaccines and in gene therapy. The polypeptides and
CC   polynucleotides encoding them, antigen presenting cells which express
CC   the polypeptides, antibodies against the polypeptides and vaccines
CC   comprising them can be used for inhibiting the development of prostate
CC   cancer in a patient. The polypeptides can be used to generate antibodies
CC   or anti-idiotypic antibodies for passive immuno therapy. A portion of
CC   the polynucleotides encoding the polypeptides can be used as a probe or
CC   to modulate the expression of the polypeptides. AAA06241 to AAA06691 and
CC   AAY82000 to AAY82020 represent sequences used in the exemplification of
CC   the present invention.
XX
SQ   Sequence   553 AA;


  Query Match             53.6%;  Score 15;  DB 21;  Length 553;
  Best Local Similarity   100.0%;  Pred. No. 8.8e-07;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 DFVGEGLYQGVPRAE 15
           |||||||||||||||
Db     294 dfvgeglyqgvprae 308


RESULT   8
AAU69763
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: June 26, 2002, 12:36:27 ; Search time 67.61 Seconds
(without alignments)
199.783 Million cell updates/sec

Title: US-09-838-785-2
Perfect score: 2861
Sequence: 1 MVQRLWVSRLLRHRKAQLLL.....AIYFATQVVFDKSDLAKYSA 553

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2861	100.0	553	4	US-09-020-956-113	Sequence 113, App
2	2861	100.0	553	4	US-09-030-607-113	Sequence 113, App
3	2861	100.0	553	4	US-09-439-313-113	Sequence 113, App
4	1287	45.0	255	4	US-09-071-710-36	Sequence 36, Appl
5	1287	45.0	255	4	US-09-525-397-36	Sequence 36, Appl
6	452	15.8	84	4	US-09-439-313-571	Sequence 571, App
7	330	11.5	516	1	US-08-356-340-4	Sequence 4, Appli
8	330	11.5	516	2	US-08-786-555-4	Sequence 4, Appli
9	324.5	11.3	525	1	US-08-356-340-2	Sequence 2, Appli
10	324.5	11.3	525	2	US-08-786-555-2	Sequence 2, Appli
11	312	10.9	58	4	US-09-439-313-547	Sequence 547, App
12	304	10.6	56	4	US-09-439-313-564	Sequence 564, App
13	243	8.5	44	4	US-09-071-710-37	Sequence 37, Appl
14	243	8.5	44	4	US-09-525-397-37	Sequence 37, Appl
15	157	5.5	29	4	US-09-439-313-546	Sequence 546, App
16	146	5.1	27	4	US-09-071-710-39	Sequence 39, Appl
17	146	5.1	27	4	US-09-525-397-39	Sequence 39, Appl
18	146	5.1	27	4	US-09-439-313-566	Sequence 566, App
19	137	4.8	25	4	US-09-439-313-520	Sequence 520, App
20	135	4.7	27	4	US-09-071-710-38	Sequence 38, Appl
21	135	4.7	27	4	US-09-525-397-38	Sequence 38, Appl
22	120	4.2	21	4	US-09-439-313-496	Sequence 496, App
23	120	4.2	21	4	US-09-439-313-521	Sequence 521, App
24	119	4.2	22	4	US-09-439-313-565	Sequence 565, App
25	115	4.0	22	4	US-09-439-313-559	Sequence 559, App
26	114	4.0	22	4	US-09-439-313-557	Sequence 557, App
27	114	4.0	22	4	US-09-439-313-561	Sequence 561, App
28	112	3.9	20	4	US-09-439-313-495	Sequence 495, App
29	110	3.8	427	4	US-08-900-230-4	Sequence 4, Appli
30	109.5	3.8	455	1	US-08-035-928-2	Sequence 2, Appli
31	109	3.8	22	4	US-09-439-313-555	Sequence 555, App
32	107	3.7	20	4	US-09-439-313-493	Sequence 493, App
33	107	3.7	20	4	US-09-439-313-497	Sequence 497, App
34	106.5	3.7	436	6	5432081-10	Patent No. 5432081
35	106	3.7	22	4	US-09-439-313-569	Sequence 569, App
36	106	3.7	427	4	US-09-199-737-4	Sequence 4, Appli
37	105	3.7	22	4	US-09-439-313-553	Sequence 553, App
38	105	3.7	22	4	US-09-439-313-572	Sequence 572, App
39	101	3.5	18	4	US-09-439-313-548	Sequence 548, App
40	100	3.5	20	4	US-09-439-313-490	Sequence 490, App
41	100	3.5	20	4	US-09-439-313-499	Sequence 499, App
42	99	3.5	20	4	US-09-439-313-563	Sequence 563, App
43	99	3.5	20	4	US-09-439-313-570	Sequence 570, App
44	98	3.4	20	4	US-09-439-313-498	Sequence 498, App
45	98	3.4	713	3	US-09-335-409-11	Sequence 11, Appl

ALIGNMENTS

```
RESULT 1
US-09-020-956-113
; Sequence 113, Application US/09020956
; Patent No. 6261562
;  GENERAL INFORMATION:
;    APPLICANT:  Xu, Jiangchun
;    APPLICANT:  Dillin, Davin C.
;    TITLE OF INVENTION:  COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE CANCER AND METHODS
;    NUMBER OF SEQUENCES:  178
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  WA
;      COUNTRY:  USA
;      ZIP:  98104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/020,956
;      FILING DATE:  09-FEB-1998
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Maki, David J.
;      REGISTRATION NUMBER:  31,392
;      REFERENCE/DOCKET NUMBER:  210121.427C2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;  INFORMATION FOR SEQ ID NO:  113:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  553 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
US-09-020-956-113


  Query Match           100.0%;  Score 2861;  DB 4;  Length 553;
  Best Local Similarity  100.0%;  Pred. No. 3.7e-271;
  Matches  553;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 MVQRLWVSRLLRHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db      1 MVQRLWVSRLLLRHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIG 60

Qy     61 PVLGLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDPRPL 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 PVLGLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDPRPL 120

Qy    121 ELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPA 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 ELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPA 180

Qy    181 IDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPH 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 IDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPH 240

Qy    241 CCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGL 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 CCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGL 300

Qy    301 YQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVA 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 YQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVA 360

Qy    361 AFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 AFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420

Qy    421 ASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 ASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480

Qy    481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFATQ 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFATQ 540

Qy    541 VVFDKSDLAKYSA 553
          |||||||||||||
Db    541 VVFDKSDLAKYSA 553


RESULT   2
US-09-030-607-113
; Sequence 113, Application US/09030607
; Patent No. 6262245
;  GENERAL INFORMATION:
;    APPLICANT:  Xu, Jiangchun
;    APPLICANT:  Dillon, Davin C.
;    TITLE OF INVENTION:  COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE CANCER AND METHODS FO
;    NUMBER OF SEQUENCES:  224
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  WA
;      COUNTRY:  USA
;      ZIP:  98104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/030,607
;      FILING DATE:  25-FEB-1998
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Maki, David J.
;      REGISTRATION NUMBER:  31,392
;      REFERENCE/DOCKET NUMBER:  210121.427C3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;  INFORMATION FOR SEQ ID NO:  113:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  553 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
US-09-030-607-113


  Query Match           100.0%;  Score 2861;  DB 4;  Length 553;
  Best Local Similarity  100.0%;  Pred. No. 3.7e-271;
  Matches  553;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 MVQRLWVSRLLLRHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIG 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MVQRLWVSRLLLRHRKAQLLLVNLLTFGLEVCLAAGITYVPPLLLEVGVEEKFMTMVLGIG 60

Qy     61 PVLGLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDPRPL 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 PVLGLVCVPLLGSASDHWRGRYGRRRPFIWALSLGILLSLFLIPRAGWLAGLLCPDPRPL 120

Qy    121 ELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPA 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 ELALLILGVGLLDFCGQVCFTPLEALLSDLFRDPDHCRQAYSVYAFMISLGGCLGYLLPA 180

Qy    181 IDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPH 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 IDWDTSALAPYLGTQEECLFGLLTLIFLTCVAATLLVAEEAALGPTEPAEGLSAPSLSPH 240

Qy    241 CCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGL 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 CCPCRARLAFRNLGALLPRLHQLCCRMPRTLRRLFVAELCSWMALMTFTLFYTDFVGEGL 300

Qy    301 YQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVA 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 YQGVPRAEPGTEARRHYDEGVRMGSLGLFLQCAISLVFSLVMDRLVQRFGTRAVYLASVA 360

Qy    361 AFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 AFPVAAGATCLSHSVAVVTASAALTGFTFSALQILPYTLASLYHREKQVFLPKYRGDTGG 420

Qy    421 ASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 ASSEDSLMTSFLPGPKPGAPFPNGHVGAGGSGLLPPPPALCGASACDVSVRVVVGEPTEA 480

Qy    481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFATQ 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 RVVPGRGICLDLAILDSAFLLSQVAPSLFMGSIVQLSQSVTAYMVSAAGLGLVAIYFATQ 540

Qy    541 VVFDKSDLAKYSA 553
          |||||||||||||
Db    541 VVFDKSDLAKYSA 553


RESULT   3
US-09-439-313-113
; Sequence 113, Application US/09439313
; Patent No. 6329505
; GENERAL INFORMATION:
;  APPLICANT: Xu, Jiangchun
;  APPLICANT:  Dillon, Davin C.
;  APPLICANT:  Mitcham, Jennifer L.
;  APPLICANT:  Harlocker, Susan Louise
;  APPLICANT:  Jiang Yuqui
;  APPLICANT:  Reed, Steven G.
;  APPLICANT:  Kalos, Michael
;  APPLICANT:  Fanger, Gary
;  APPLICANT:  Retter, Mark
;  APPLICANT:  Solk, John
;  APPLICANT:  Day, Craig
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
;  TITLE OF INVENTION:  DIAGNOSIS OF PROSTATE CANCER
```

ID AAU69763 standard; Protein; 553 AA.
XX
AC AAU69763;
XX
DT 30-JAN-2002 (first entry)
XX
DE Human prostate cDNA encoded protein #3.
XX
KW Human; prostate cancer; cytostatic; immunostimulant; tumour; immunogen.
XX
OS Homo sapiens.
XX
PN WO200173032-A2.
XX
PD 04-OCT-2001.
XX
PF 27-MAR-2001; 2001WO-US09919.
XX
PR 27-MAR-2000; 2000US-0536857.
PR 09-MAY-2000; 2000US-0568100.
PR 12-MAY-2000; 2000US-0570737.
PR 13-JUN-2000; 2000US-0593793.
PR 27-JUN-2000; 2000US-0605783.
PR 10-AUG-2000; 2000US-0636215.
PR 29-AUG-2000; 2000US-0651236.
PR 06-SEP-2000; 2000US-0657279.
PR 02-OCT-2000; 2000US-0679426.
PR 10-OCT-2000; 2000US-0685166.
XX
PA (CORI-) CORIXA CORP.
XX
PI Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Kalos MD;
PI Fanger GR, Retter MW, Stolk JA, Day CH, Vedvick TS, Carter D;
PI Li SX, Wang A, Skeiky YAW, Hepler WT, Henderson RA;
XX
DR WPI; 2001-639232/73.
DR N-PSDB; AAS63557.
XX
PT New human prostate-specific polypeptides and polynucleotides useful for
PT the diagnosis and treatment of cancer, especially prostate cancer -
XX
PS Claim 2; Page 269-270; 579pp; English.
XX
CC The invention relates to isolated prostate-specific
CC polynucleotides, polypeptides, fusion proteins of the polypeptides,
CC antibodies raised against the polypeptides (or antigenic epitopes
CC derived from them) and antigen-presenting cells expressing the
CC polypeptides. The antibodies are useful for detecting the presence of
CC cancer, especially prostate cancer. The polypeptides, polynucleotides and
CC the antigen-presenting cells are useful for stimulating and/or expanding
CC T cells specific for a tumour protein, and for inhibiting the development
CC of cancer especially prostate cancer. Compositions comprising the
CC polynucleotide and/or polypeptide are useful for stimulating an immune
CC response, and for treating cancer. The oligonucleotide is useful for
CC detecting cancer. The present sequence is a prostate specific
CC polypeptide of the invention.
XX
SQ Sequence 553 AA;

Query Match 53.6%; Score 15; DB 22; Length 553;
Best Local Similarity 100.0%; Pred. No. 8.8e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DFVGEGLYQGVPRAE 15
||||||||||||||||
Db 294 dfvgeglyqgvprae 308

RESULT 9
AAU04961
ID AAU04961 standard; Protein; 553 AA.
XX
AC AAU04961;
XX
DT 24-OCT-2001 (first entry)
XX
DE Human prostate tumour protein L1-12.
XX
KW Human; prostate tumour protein; prostate cancer.
XX
OS Homo sapiens.
XX
PN US6262245-B1.
XX
PD 17-JUL-2001.
XX
PF 25-FEB-1998; 98US-0030607.
XX
PR 25-FEB-1997; 97US-0806099.
PR 01-AUG-1997; 97US-0904804.
PR 09-FEB-1998; 98US-0020956.
XX
PA (CORI-) CORIXA CORP.
XX
PI Xu J, Dillon DC;
XX
DR WPI; 2001-440862/47.
DR N-PSDB; AAS10108.
XX
PT Novel polynucleotide encoding polypeptide comprising a portion of
PT prostate tumour protein useful for inhibiting development of prostate
PT cancer or for treating prostate cancer in a patient -
XX
PS Example 1; Column 125-127; 105pp; English.
XX
CC The sequence is a partial prostate tumour protein, encoded by a prostate
CC tumour specific cDNA. The DNA is useful for inhibiting the development
CC of prostate cancer or for treating prostate cancer in a patient.
XX
SQ Sequence 553 AA;

Query Match 53.6%; Score 15; DB 22; Length 553;
Best Local Similarity 100.0%; Pred. No. 8.8e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DFVGEGLYQGVPRAE 15
||||||||||||||||
Db 294 dfvgeglyqgvprae 308

RESULT 10
AAM01117
ID AAM01117 standard; Protein; 553 AA.
XX
AC AAM01117;
XX
DT 04-OCT-2001 (first entry)
XX
DE Human prostate-specific amino acid sequence L1-12.
XX
KW Human; prostate cancer; prostate-specific; diagnosis; vaccine;
KW cytostatic; gene therapy; metastasis.
XX
OS Homo sapiens.
XX
PN WO200151633-A2.
XX
PD 19-JUL-2001.
XX
PF 16-JAN-2001; 2001WO-US01574.
XX
PR 14-JAN-2000; 2000US-0483672.
XX
PA (CORI-) CORIXA CORP.

```
XX
PI   Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Reed SG;
PI   Kalos MD, Fanger GR, Day CH, Retter MW, Stolk JA, Skeiky YAW;
PI   Wang A, Meagher MJ;
XX
DR   WPI; 2001-425873/45.
XX
PT   New polynucleotide encoding a prostate-specific protein, for
PT   diagnosing, monitoring and treating prostate cancer in a patient and
PT   for use in vaccines -
XX
PS   Claim 2; Page 267-268; 543pp; English.
XX
CC   The present invention describes polynucleotide sequences (I) which encode
CC   prostate-specific proteins (II). (I) and (II) have cytostatic activity,
CC   and can be used in vaccine production and gene therapy. (I), (II),
CC   antibodies to (II), fusion proteins comprising (II), and isolated
CC   T cells prepared using (I) or (II) are used treat cancer in a patient.
CC   (I) and the antibodies are also used in the detection of cancer in a
CC   patient. The cancer that is diagnosed or treated is particularly
CC   prostate cancer. (I) and (II) can be used in vaccines. The antibodies or
CC   (I) can be used for monitoring the progression of cancer in a patient.
CC   (I) and (II) can also be used to improve diagnostic and therapeutic
CC   methods for prostate cancer. They can indicate the level of metastasis
CC   as well as the prostate volume. AAH93357 to AAH93944 and AAM01115 to
CC   AAM01318 represent polynucleotide and amino acid sequences used in the
CC   exemplification of the present invention.
XX
SQ   Sequence   553 AA;


  Query Match             53.6%;  Score 15;  DB 22;  Length 553;
  Best Local Similarity   100.0%;  Pred. No. 8.8e-07;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 DFVGEGLYQGVPRAE 15
           |||||||||||||||
Db     294 dfvgeglyqgvprae 308


RESULT  11
AAG99002
ID   AAG99002 standard; Protein; 553 AA.
XX
AC   AAG99002;
XX
DT   25-SEP-2001  (first entry)
XX
DE   Human prostate-specific amino acid sequence L1-12/P501S.
XX
KW   Human; prostate cancer; therapy; diagnosis; cat eye syndrome;
KW   chromosome 22q11.2; prostate-specific protein; chromosome 1;
KW   prostate specific antigen; PSA.
XX
OS   Homo sapiens.
XX
PN   WO200134802-A2.
XX
PD   17-MAY-2001.
XX
PF   09-NOV-2000; 2000WO-US30904.
XX
PR   12-NOV-1999;   99US-0439313.
PR   18-NOV-1999;   99US-0443686.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Reed SG;
PI   Kalos MD, Retter MW, Stolk JA, Day CH, Skeiky YAW, Wang A;
XX
DR   WPI; 2001-308785/32.
XX
PT   Isolated polypeptide comprising at least an immunogenic portion of a
PT   prostate-specific protein, useful in the diagnosis and therapy of
PT   prostate cancer -
XX
PS   Claim 3; Page 167-168; 325pp; English.
XX
CC   The present invention describes an isolated polypeptide (P1) comprising
CC   at least an immunogenic portion of a prostate-specific protein, or its
CC   variant. Also described are polynucleotides (N1) encoding (P1). (P1) and
CC   (N1) have cytostatic activity and can be used in vaccine production.
CC   The polypeptides, nucleic acids and antibodies from the present
CC   invention are useful in the diagnosis and therapy of prostate cancer.
CC   Prostate specific genes P704P, P712P, P774P, P775P and B305D are located
CC   in a genomic region on chromosome 22q11.2 known as the Cat Eye Syndrome
CC   region. Prostate specific antigen (PSA) P501S was located on
CC   chromosome 1. AAH84671 to AAH85143 and AAG99000 to AAG99077 represent
CC   polynucleotide and polypeptide sequences used in the exemplification
CC   of the present invention.
XX
SQ   Sequence   553 AA;


  Query Match             53.6%;  Score 15;  DB 22;  Length 553;
  Best Local Similarity   100.0%;  Pred. No. 8.8e-07;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 DFVGEGLYQGVPRAE 15
           |||||||||||||||
Db     294 dfvgeglyqgvprae 308


RESULT  12
AAG62150
ID   AAG62150 standard; Protein; 553 AA.
XX
AC   AAG62150;
XX
DT   06-JUL-2001  (first entry)
XX
DE   Human P501S inventive antigen SEQ ID NO: 333.
XX
KW   Human; mouse; immunotherapy; cancer; leukaemia; WT1; Wilm's tumour gene;
KW   chromosome 11p13; zinc finger transcription factor.
XX
OS   Homo sapiens.
XX
PN   WO200125273-A2.
XX
PD   12-APR-2001.
XX
PF   04-OCT-2000; 2000WO-US27465.
XX
PR   04-OCT-1999;   99US-0157459.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Skeiky YAW, Xu J, Cheever MA, Reed SG;
XX
DR   WPI; 2001-328324/34.
XX
PT   Polypeptide comprising part of the Wilm's Tumour gene product sequence is
PT   used in the diagnosis and treatment of malignant diseases e.g. leukemia
PT   and cancer associated with WTI -
XX
PS   Disclosure; Page 212-213; 228pp; English.
XX
CC   The present invention describes compositions comprising peptides derived
CC   from the Wilm's tumour protein WT1 and methods for their use in treating
CC   malignant diseases. Peptides derived from both the murine and human WT1
CC   proteins are provided. The human WT1 gene is found on chromosome 11p13,
CC   and the protein was shown to be a zinc finger transcription factor. The
CC   immunogenic peptides of the invention are particularly useful in the
CC   diagnosis and treatment of cancer and leukaemia. The present sequence is
CC   a polypeptide described in the exemplification of the invention.
```

XX
SQ Sequence 553 AA;

Query Match 53.6%; Score 15; DB 22; Length 553;
Best Local Similarity 100.0%; Pred. No. 8.8e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DFVGEGLYQGVPRAE 15
||||||||||||||||
Db 294 dfvgeglyqgvprae 308

RESULT 13
AAB74800
ID AAB74800 standard; Protein; 553 AA.
XX
AC AAB74800;
XX
DT 14-JUN-2001 (first entry)
XX
DE Prostate tumour antigen predicted amino acid sequence for L1-12.
XX
KW Human; prostate tumour antigen; prostate tumour; therapy; diagnosis;
KW prostate cancer; immunogenic; cytostatic; vaccine.
XX
OS Homo sapiens.
XX
PN WO200125272-A2.
XX
PD 12-APR-2001.
XX
PF 04-OCT-2000; 2000WO-US27464.
XX
PR 04-OCT-1999; 99US-0157455.
XX
PA (CORI-) CORIXA CORP.
XX
PI Xu J, Skeiky YAW, Reed SG, Cheever MA;
XX
DR WPI; 2001-245062/25.
DR N-PSDB; AAH02530.
XX
PT Prostate specific protein and its encoding polynucleotide, useful for
PT the treatment and diagnosis of prostate cancer -
XX
PS Claim 3; Page 157-158; 276pp; English.
XX
CC The present invention describes an isolated polypeptide (I) comprising
CC at least an immunogenic portion of a prostate tumour antigen protein or
CC its variant. (I) have cytostatic activity and can be used in vaccine
CC production. (I), prostate tumour antigen polynucleotides, an antigen
CC presenting cell (APC e.g. a dendritic cell) that expresses (I), and a
CC pharmaceutical composition containing (I) are useful for inhibiting the
CC development of cancer in a patient. Antibodies specific for prostate
CC specific proteins and oligonucleotides that hybridise to a
CC polynucleotide that encodes a prostate specific protein are useful
CC for detecting the presence or absence of a cancer or monitoring the
CC progression the progression of a cancer, especially prostate cancer.
CC AAH02422 to AAH2872, AAB74798 to AAB74821 and AAB74830 are sequences
CC used in the exemplification of the present invention.
XX
SQ Sequence 553 AA;

Query Match 53.6%; Score 15; DB 22; Length 553;
Best Local Similarity 100.0%; Pred. No. 8.8e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DFVGEGLYQGVPRAE 15
||||||||||||||||
Db 294 dfvgeglyqgvprae 308

RESULT 14
AAM01318
ID AAM01318 standard; Protein; 595 AA.
XX
AC AAM01318;
XX
DT 04-OCT-2001 (first entry)
XX
DE Alpha prepro-P501S recombinant protein amino acid sequence.
XX
KW Human; prostate cancer; prostate-specific; diagnosis; vaccine;
KW cytostatic; gene therapy; metastasis.
XX
OS Homo sapiens.
XX
PN WO200151633-A2.
XX
PD 19-JUL-2001.
XX
PF 16-JAN-2001; 2001WO-US01574.
XX
PR 14-JAN-2000; 2000US-0483672.
XX
PA (CORI-) CORIXA CORP.
XX
PI Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Reed SG;
PI Kalos MD, Fanger GR, Day CH, Retter MW, Stolk JA, Skeiky YAW;
PI Wang A, Meagher MJ;
XX
DR WPI; 2001-425873/45.
XX
PT New polynucleotide encoding a prostate-specific protein, for
PT diagnosing, monitoring and treating prostate cancer in a patient and
PT for use in vaccines -
XX
PS Example 17; Page 541-543; 543pp; English.
XX
CC The present invention describes polynucleotide sequences (I) which encode
CC prostate-specific proteins (II). (I) and (II) have cytostatic activity,
CC and can be used in vaccine production and gene therapy. (I), (II),
CC antibodies to (II), fusion proteins comprising (II), and isolated
CC T cells prepared using (I) or (II) are used treat cancer in a patient.
CC (I) and the antibodies are also used in the detection of cancer in a
CC patient. The cancer that is diagnosed or treated is particularly
CC prostate cancer. (I) and (II) can be used in vaccines. The antibodies or
CC (I) can be used for monitoring the progression of cancer in a patient.
CC (I) and (II) can also be used to improve diagnostic and therapeutic
CC methods for prostate cancer. They can indicate the level of metastasis
CC as well as the prostate volume. AAH93357 to AAH93944 and AAM01115 to
CC AAM01318 represent polynucleotide and amino acid sequences used in the
CC exemplification of the present invention.
XX
SQ Sequence 595 AA;

Query Match 53.6%; Score 15; DB 22; Length 595;
Best Local Similarity 100.0%; Pred. No. 9.3e-07;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DFVGEGLYQGVPRAE 15
||||||||||||||||
Db 328 dfvgeglyqgvprae 342

RESULT 15
AAB74830
ID AAB74830 standard; Protein; 1079 AA.
XX
AC AAB74830;
XX
DT 14-JUN-2001 (first entry)
XX

DE Prostate tumour antigen amino acid sequence for a fusion protein.
XX
KW Human; prostate tumour antigen; prostate tumour; therapy; diagnosis;
KW prostate cancer; immunogenic; cytostatic; vaccine.
XX
OS Homo sapiens.
XX
PN WO200125272-A2.
XX
PD 12-APR-2001.
XX
PF 04-OCT-2000; 2000WO-US27464.
XX
PR 04-OCT-1999; 99US-0157455.
XX
PA (CORI-) CORIXA CORP.
XX
PI Xu J, Skeiky YAW, Reed SG, Cheever MA;
XX
DR WPI; 2001-245062/25.
XX
PT Prostate specific protein and its encoding polynucleotide, useful for
PT the treatment and diagnosis of prostate cancer -
XX
PS Disclosure; Page 272-276; 276pp; English.
XX
CC The present invention describes an isolated polypeptide (I) comprising
CC at least an immunogenic portion of a prostate tumour antigen protein or
CC its variant. (I) have cytostatic activity and can be used in vaccine
CC production. (I), prostate tumour antigen polynucleotides, an antigen
CC presenting cell (APC e.g. a dendritic cell) that expresses (I), and a
CC pharmaceutical composition containing (I) are useful for inhibiting the
CC development of cancer in a patient. Antibodies specific for prostate
CC specific proteins and oligonucleotides that hybridise to a
CC polynucleotide that encodes a prostate specific protein are useful
CC for detecting the presence or absence of a cancer or monitoring the
CC progression the progression of a cancer, especially prostate cancer.
CC AAH02422 to AAH2872, AAB74798 to AAB74821 and AAB74830 are sequences
CC used in the exemplification of the present invention.
XX
SQ Sequence 1079 AA;

Query Match 53.6%; Score 15; DB 22; Length 1079;
Best Local Similarity 100.0%; Pred. No. 1.5e-06;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DFVGEGLYQGVPRAE 15
||||||||||||||||
Db 820 dfvgeglyqgvprae 834

Search completed: June 26, 2002, 13:49:00
Job time: 353 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: June 26, 2002, 13:44:22 ; Search time 14.83 Seconds
(without alignments)
46.117 Million cell updates/sec

Title: US-09-838-785-25
Perfect score: 28
Sequence: 1 DFVGEGLYQGVPRAEGTEARRHYDEGVR 28

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 231628 seqs, 24425594 residues

Word size : 0

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	15	53.6	553	4	US-09-020-956-113	Sequence 113, App
2	15	53.6	553	4	US-09-030-607-113	Sequence 113, App
3	15	53.6	553	4	US-09-439-313-113	Sequence 113, App
4	14	50.0	29	4	US-09-439-313-546	Sequence 546, App
5	13	46.4	25	4	US-09-439-313-520	Sequence 520, App
6	13	46.4	27	4	US-09-071-710-39	Sequence 39, Appl
7	13	46.4	27	4	US-09-525-397-39	Sequence 39, Appl
8	13	46.4	27	4	US-09-439-313-566	Sequence 566, App
9	13	46.4	255	4	US-09-071-710-36	Sequence 36, Appl
10	13	46.4	255	4	US-09-525-397-36	Sequence 36, Appl
11	11	39.3	15	4	US-09-439-313-518	Sequence 518, App
12	7	25.0	1958	1	US-07-945-283-2	Sequence 2, Appli
13	6	21.4	17	4	US-09-439-313-519	Sequence 519, App
14	6	21.4	327	4	US-09-290-640-66	Sequence 66, Appl
15	6	21.4	525	6	5183745-4	Patent No. 5183745
16	6	21.4	1489	6	5183745-2	Patent No. 5183745
17	6	21.4	1705	4	US-08-669-785-4	Sequence 4, Appli
18	6	21.4	1706	4	US-08-669-785-2	Sequence 2, Appli
19	6	21.4	1794	6	5183745-6	Patent No. 5183745
20	5	17.9	8	4	US-09-420-211-12	Sequence 12, Appl
21	5	17.9	9	2	US-08-951-924A-3	Sequence 3, Appli
22	5	17.9	11	3	US-08-592-500-42	Sequence 42, Appl
23	5	17.9	11	3	US-08-195-006-42	Sequence 42, Appl
24	5	17.9	11	5	PCT-US94-07644A-42	Sequence 42, Appl
25	5	17.9	14	1	US-08-277-299-1	Sequence 1, Appli
26	5	17.9	14	2	US-08-880-671-1	Sequence 1, Appli
27	5	17.9	14	4	US-08-853-910-5	Sequence 5, Appli
28	5	17.9	15	1	US-07-969-336-1	Sequence 1, Appli
29	5	17.9	15	2	US-08-815-953-1	Sequence 1, Appli
30	5	17.9	17	2	US-08-472-659-4	Sequence 4, Appli
31	5	17.9	17	2	US-08-474-661-4	Sequence 4, Appli
32	5	17.9	17	2	US-08-611-977-4	Sequence 4, Appli
33	5	17.9	18	1	US-07-920-281C-27	Sequence 27, Appl
34	5	17.9	18	4	US-08-466-277-27	Sequence 27, Appl
35	5	17.9	19	4	US-08-975-040-8	Sequence 8, Appli
36	5	17.9	26	2	US-08-563-892A-1	Sequence 1, Appli
37	5	17.9	26	2	US-08-563-892A-20	Sequence 20, Appl
38	5	17.9	26	2	US-08-563-892A-21	Sequence 21, Appl
39	5	17.9	26	2	US-08-563-892A-22	Sequence 22, Appl
40	5	17.9	38	2	US-08-284-391B-50	Sequence 50, Appl
41	5	17.9	38	4	US-09-218-950-50	Sequence 50, Appl
42	5	17.9	53	2	US-08-563-892A-29	Sequence 29, Appl
43	5	17.9	61	4	US-09-537-357-13	Sequence 13, Appl
44	5	17.9	61	4	US-09-537-357-14	Sequence 14, Appl
45	5	17.9	64	2	US-08-563-892A-27	Sequence 27, Appl

ALIGNMENTS

```
RESULT   1
US-09-020-956-113
; Sequence 113, Application US/09020956
; Patent No. 6261562
;  GENERAL INFORMATION:
;    APPLICANT:  Xu, Jiangchun
;    APPLICANT:  Dillin, Davin C.
;    TITLE OF INVENTION:  COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE CANCER AND METHODS
;    NUMBER OF SEQUENCES:  178
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  WA
;      COUNTRY:  USA
;      ZIP:  98104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/020,956
;      FILING DATE:  09-FEB-1998
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Maki, David J.
;      REGISTRATION NUMBER:  31,392
;      REFERENCE/DOCKET NUMBER:  210121.427C2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;  INFORMATION FOR SEQ ID NO:  113:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  553 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
US-09-020-956-113


  Query Match             53.6%;  Score 15;  DB 4;  Length 553;
  Best Local Similarity   100.0%;  Pred. No. 2.6e-07;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 DFVGEGLYQGVPRAE 15
           |||||||||||||||
```

```
Db    294 DFVGEGLYQGVPRAE 308


RESULT   2
US-09-030-607-113
; Sequence 113, Application US/09030607
; Patent No. 6262245
;  GENERAL INFORMATION:
;    APPLICANT:  Xu, Jiangchun
;    APPLICANT:  Dillon, Davin C.
;    TITLE OF INVENTION:  COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE CANCER AND METHODS FO
;    NUMBER OF SEQUENCES:  224
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  WA
;      COUNTRY:  USA
;      ZIP:  98104
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/030,607
;      FILING DATE:  25-FEB-1998
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Maki, David J.
;      REGISTRATION NUMBER:  31,392
;      REFERENCE/DOCKET NUMBER:  210121.427C3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;  INFORMATION FOR SEQ ID NO:  113:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  553 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
US-09-030-607-113


  Query Match             53.6%;  Score 15;  DB 4;  Length 553;
  Best Local Similarity   100.0%;  Pred. No. 2.6e-07;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 DFVGEGLYQGVPRAE 15
           |||||||||||||||
Db     294 DFVGEGLYQGVPRAE 308


RESULT   3
US-09-439-313-113
; Sequence 113, Application US/09439313
; Patent No. 6329505
; GENERAL INFORMATION:
;  APPLICANT: Xu, Jiangchun
;  APPLICANT:  Dillon, Davin C.
;  APPLICANT:  Mitcham, Jennifer L.
;  APPLICANT:  Harlocker, Susan Louise
;  APPLICANT:  Jiang Yuqui
;  APPLICANT:  Reed, Steven G.
;  APPLICANT:  Kalos, Michael
;  APPLICANT:  Fanger, Gary
;  APPLICANT:  Retter, Mark
;  APPLICANT:  Solk, John
;  APPLICANT:  Day, Craig
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
;  TITLE OF INVENTION:  DIAGNOSIS OF PROSTATE CANCER
;  FILE REFERENCE: 210121.427C9
;  CURRENT APPLICATION NUMBER: US/09/439,313
;  CURRENT FILING DATE:  1999-11-12
;  NUMBER OF SEQ ID NOS: 575
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 113
;   LENGTH: 553
;   TYPE: PRT
;   ORGANISM: Homo sapien
US-09-439-313-113


  Query Match             53.6%;  Score 15;  DB 4;  Length 553;
  Best Local Similarity   100.0%;  Pred. No. 2.6e-07;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 DFVGEGLYQGVPRAE 15
           |||||||||||||||
Db     294 DFVGEGLYQGVPRAE 308


RESULT   4
US-09-439-313-546
; Sequence 546, Application US/09439313
; Patent No. 6329505
; GENERAL INFORMATION:
;  APPLICANT: Xu, Jiangchun
;  APPLICANT:  Dillon, Davin C.
;  APPLICANT:  Mitcham, Jennifer L.
;  APPLICANT:  Harlocker, Susan Louise
;  APPLICANT:  Jiang Yuqui
;  APPLICANT:  Reed, Steven G.
;  APPLICANT:  Kalos, Michael
;  APPLICANT:  Fanger, Gary
;  APPLICANT:  Retter, Mark
;  APPLICANT:  Solk, John
;  APPLICANT:  Day, Craig
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
;  TITLE OF INVENTION:  DIAGNOSIS OF PROSTATE CANCER
;  FILE REFERENCE: 210121.427C9
;  CURRENT APPLICATION NUMBER: US/09/439,313
;  CURRENT FILING DATE:  1999-11-12
;  NUMBER OF SEQ ID NOS: 575
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 546
;   LENGTH: 29
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-439-313-546


  Query Match             50.0%;  Score 14;  DB 4;  Length 29;
  Best Local Similarity   100.0%;  Pred. No. 2.1e-07;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 FVGEGLYQGVPRAE 15
           ||||||||||||||
Db       1 FVGEGLYQGVPRAE 14


RESULT   5
US-09-439-313-520
; Sequence 520, Application US/09439313
; Patent No. 6329505
; GENERAL INFORMATION:
;  APPLICANT: Xu, Jiangchun
;  APPLICANT:  Dillon, Davin C.
;  APPLICANT:  Mitcham, Jennifer L.
;  APPLICANT:  Harlocker, Susan Louise
;  APPLICANT:  Jiang Yuqui
;  APPLICANT:  Reed, Steven G.
;  APPLICANT:  Kalos, Michael
```

```
;  APPLICANT:  Fanger, Gary
;  APPLICANT:  Retter, Mark
;  APPLICANT:  Solk, John
;  APPLICANT:  Day, Craig
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
;  TITLE OF INVENTION:  DIAGNOSIS OF PROSTATE CANCER
;  FILE REFERENCE: 210121.427C9
;  CURRENT APPLICATION NUMBER: US/09/439,313
;  CURRENT FILING DATE:  1999-11-12
;  NUMBER OF SEQ ID NOS: 575
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 520
;   LENGTH: 25
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Made in a lab
US-09-439-313-520


  Query Match             46.4%;  Score 13;  DB 4;  Length 25;
  Best Local Similarity   100.0%;  Pred. No. 1.7e-06;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 VGEGLYQGVPRAE 15
           |||||||||||||
Db       1 VGEGLYQGVPRAE 13


RESULT   6
US-09-071-710-39
; Sequence 39, Application US/09071710
; Patent No. 6130043
;  GENERAL INFORMATION:
;    APPLICANT:  BILLING-MEDEL, PATRICIA
;    APPLICANT:  COHEN, MAURICE
;    APPLICANT:  COLPITTS, TRACEY L.
;    APPLICANT:  FRIEDMAN, PAULA N.
;    APPLICANT:  GORDON, JULIAN
;    APPLICANT:  GRANADOS, EDWARD N.
;    APPLICANT:  HODGES, STEVEN C.
;    APPLICANT:  KLASS, MICHAEL R.
;    APPLICANT:  KRATOCHVIL, JON D.
;    APPLICANT:  ROBERTS-RAPP, LISA
;    APPLICANT:  RUSSELL, JOHN C.
;    APPLICANT:  STROUPE, STEPHEN D.
;    TITLE OF INVENTION:  REAGENTS AND METHODS USEFUL
;    TITLE OF INVENTION:  FOR DETECTING DISEASES OF THE PROSTATE
;    NUMBER OF SEQUENCES:  41
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Abbott Laboratories
;      STREET:  100 Abbott Park Road
;      CITY:  Abbott Park
;      STATE:  IL
;      COUNTRY:  USA
;      ZIP:  60064-3500
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/071,710
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/850,713
;      FILING DATE:  02-MAY-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Becker, Cheryl L.
;      REGISTRATION NUMBER:  35,441
;      REFERENCE/DOCKET NUMBER:  6083.US.P1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  847/935-1729
;      TELEFAX:  847/938-2623
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  39:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  27 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  No. 6130043e
US-09-071-710-39


  Query Match             46.4%;  Score 13;  DB 4;  Length 27;
  Best Local Similarity   100.0%;  Pred. No. 1.8e-06;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      16 GTEARRHYDEGVR 28
           |||||||||||||
Db      12 GTEARRHYDEGVR 24


RESULT   7
US-09-525-397-39
; Sequence 39, Application US/09525397
; Patent No. 6252047
;  GENERAL INFORMATION:
;    APPLICANT:  BILLING-MEDEL, PATRICIA
;    APPLICANT:  COHEN, MAURICE
;    APPLICANT:  COLPITTS, TRACEY L.
;    APPLICANT:  FRIEDMAN, PAULA N.
;    APPLICANT:  GORDON, JULIAN
;    APPLICANT:  GRANADOS, EDWARD N.
;    APPLICANT:  HODGES, STEVEN C.
;    APPLICANT:  KLASS, MICHAEL R.
;    APPLICANT:  KRATOCHVIL, JON D.
;    APPLICANT:  ROBERTS-RAPP, LISA
;    APPLICANT:  RUSSELL, JOHN C.
;    APPLICANT:  STROUPE, STEPHEN D.
;    TITLE OF INVENTION:  REAGENTS AND METHODS USEFUL
;    TITLE OF INVENTION:  FOR DETECTING DISEASES OF THE PROSTATE
;    NUMBER OF SEQUENCES:  41
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Abbott Laboratories
;      STREET:  100 Abbott Park Road
;      CITY:  Abbott Park
;      STATE:  IL
;      COUNTRY:  USA
;      ZIP:  60064-3500
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/525,397
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  09/071,710
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Becker, Cheryl L.
;      REGISTRATION NUMBER:  35,441
;      REFERENCE/DOCKET NUMBER:  6083.US.P1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  847/935-1729
;      TELEFAX:  847/938-2623
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  39:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  27 amino acids
;      TYPE:  amino acid
```

```
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  No. 6252047e
US-09-525-397-39


  Query Match             46.4%;  Score 13;  DB 4;  Length 27;
  Best Local Similarity   100.0%;  Pred. No. 1.8e-06;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      16 GTEARRHYDEGVR 28
           |||||||||||||
Db      12 GTEARRHYDEGVR 24


RESULT   8
US-09-439-313-566
; Sequence 566, Application US/09439313
; Patent No. 6329505
; GENERAL INFORMATION:
;  APPLICANT: Xu, Jiangchun
;  APPLICANT:  Dillon, Davin C.
;  APPLICANT:  Mitcham, Jennifer L.
;  APPLICANT:  Harlocker, Susan Louise
;  APPLICANT:  Jiang Yuqui
;  APPLICANT:  Reed, Steven G.
;  APPLICANT:  Kalos, Michael
;  APPLICANT:  Fanger, Gary
;  APPLICANT:  Retter, Mark
;  APPLICANT:  Solk, John
;  APPLICANT:  Day, Craig
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
;  TITLE OF INVENTION:  DIAGNOSIS OF PROSTATE CANCER
;  FILE REFERENCE: 210121.427C9
;  CURRENT APPLICATION NUMBER: US/09/439,313
;  CURRENT FILING DATE:  1999-11-12
;  NUMBER OF SEQ ID NOS: 575
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 566
;   LENGTH: 27
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-439-313-566


  Query Match             46.4%;  Score 13;  DB 4;  Length 27;
  Best Local Similarity   100.0%;  Pred. No. 1.8e-06;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 VGEGLYQGVPRAE 15
           |||||||||||||
Db       1 VGEGLYQGVPRAE 13


RESULT   9
US-09-071-710-36
; Sequence 36, Application US/09071710
; Patent No. 6130043
;  GENERAL INFORMATION:
;    APPLICANT:  BILLING-MEDEL, PATRICIA
;    APPLICANT:  COHEN, MAURICE
;    APPLICANT:  COLPITTS, TRACEY L.
;    APPLICANT:  FRIEDMAN, PAULA N.
;    APPLICANT:  GORDON, JULIAN
;    APPLICANT:  GRANADOS, EDWARD N.
;    APPLICANT:  HODGES, STEVEN C.
;    APPLICANT:  KLASS, MICHAEL R.
;    APPLICANT:  KRATOCHVIL, JON D.
;    APPLICANT:  ROBERTS-RAPP, LISA
;    APPLICANT:  RUSSELL, JOHN C.
;    APPLICANT:  STROUPE, STEPHEN D.
;    TITLE OF INVENTION:  REAGENTS AND METHODS USEFUL
;    TITLE OF INVENTION:  FOR DETECTING DISEASES OF THE PROSTATE
;    NUMBER OF SEQUENCES:  41
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Abbott Laboratories
;      STREET:  100 Abbott Park Road
;      CITY:  Abbott Park
;      STATE:  IL
;      COUNTRY:  USA
;      ZIP:  60064-3500
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/071,710
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/850,713
;      FILING DATE:  02-MAY-1997
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Becker, Cheryl L.
;      REGISTRATION NUMBER:  35,441
;      REFERENCE/DOCKET NUMBER:  6083.US.P1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  847/935-1729
;      TELEFAX:  847/938-2623
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  36:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  255 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  No. 6130043e
US-09-071-710-36


  Query Match             46.4%;  Score 13;  DB 4;  Length 255;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-05;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      16 GTEARRHYDEGVR 28
           |||||||||||||
Db      12 GTEARRHYDEGVR 24


RESULT  10
US-09-525-397-36
; Sequence 36, Application US/09525397
; Patent No. 6252047
;  GENERAL INFORMATION:
;    APPLICANT:  BILLING-MEDEL, PATRICIA
;    APPLICANT:  COHEN, MAURICE
;    APPLICANT:  COLPITTS, TRACEY L.
;    APPLICANT:  FRIEDMAN, PAULA N.
;    APPLICANT:  GORDON, JULIAN
;    APPLICANT:  GRANADOS, EDWARD N.
;    APPLICANT:  HODGES, STEVEN C.
;    APPLICANT:  KLASS, MICHAEL R.
;    APPLICANT:  KRATOCHVIL, JON D.
;    APPLICANT:  ROBERTS-RAPP, LISA
;    APPLICANT:  RUSSELL, JOHN C.
;    APPLICANT:  STROUPE, STEPHEN D.
;    TITLE OF INVENTION:  REAGENTS AND METHODS USEFUL
;    TITLE OF INVENTION:  FOR DETECTING DISEASES OF THE PROSTATE
;    NUMBER OF SEQUENCES:  41
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Abbott Laboratories
;      STREET:  100 Abbott Park Road
;      CITY:  Abbott Park
;      STATE:  IL
;      COUNTRY:  USA
```

```
;      ZIP:  60064-3500
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/525,397
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  09/071,710
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Becker, Cheryl L.
;      REGISTRATION NUMBER:  35,441
;      REFERENCE/DOCKET NUMBER:  6083.US.P1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  847/935-1729
;      TELEFAX:  847/938-2623
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  36:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  255 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  No. 6252047e
US-09-525-397-36


  Query Match             46.4%;  Score 13;  DB 4;  Length 255;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-05;
  Matches   13;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      16 GTEARRHYDEGVR 28
           |||||||||||||
Db      12 GTEARRHYDEGVR 24


RESULT  11
US-09-439-313-518
; Sequence 518, Application US/09439313
; Patent No. 6329505
; GENERAL INFORMATION:
;  APPLICANT: Xu, Jiangchun
;  APPLICANT:  Dillon, Davin C.
;  APPLICANT:  Mitcham, Jennifer L.
;  APPLICANT:  Harlocker, Susan Louise
;  APPLICANT:  Jiang Yuqui
;  APPLICANT:  Reed, Steven G.
;  APPLICANT:  Kalos, Michael
;  APPLICANT:  Fanger, Gary
;  APPLICANT:  Retter, Mark
;  APPLICANT:  Solk, John
;  APPLICANT:  Day, Craig
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
;  TITLE OF INVENTION:  DIAGNOSIS OF PROSTATE CANCER
;  FILE REFERENCE: 210121.427C9
;  CURRENT APPLICATION NUMBER: US/09/439,313
;  CURRENT FILING DATE:  1999-11-12
;  NUMBER OF SEQ ID NOS: 575
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 518
;    LENGTH: 15
;    TYPE: PRT
;    ORGANISM: Artificial Sequence
;    FEATURE:
;    OTHER INFORMATION: Made in a lab
US-09-439-313-518


  Query Match             39.3%;  Score 11;  DB 4;  Length 15;
  Best Local Similarity   100.0%;  Pred. No. 9.4e-05;
  Matches   11;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      16 GTEARRHYDEG 26
           ||||||||||
Db       5 GTEARRHYDEG 15


RESULT  12
US-07-945-283-2
; Sequence 2, Application US/07945283
; Patent No. 5352596
;  GENERAL INFORMATION:
;    APPLICANT:  Cheung, Andrew K.
;    APPLICANT:  Wesley, Ronald D.
;    TITLE OF INVENTION:  Pseudorabies Virus Deletion Mutants
;    TITLE OF INVENTION:  Involving The EP0 and LLT Genes
;    NUMBER OF SEQUENCES:  7
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Curtis P. Ribando
;      STREET:  1815 No. 5352596th University Street
;      CITY:  Peoria
;      STATE:  IL
;      COUNTRY:  USA
;      ZIP:  61604
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/945,283
;      FILING DATE:  19920911
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Ribando, Curtis P
;      REGISTRATION NUMBER:  27976
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  309-685-4011 ext.513
;      TELEFAX:  309-685-4128
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1958 amino acids
;      TYPE:  AMINO ACID
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-07-945-283-2


  Query Match             25.0%;  Score 7;  DB 1;  Length 1958;
  Best Local Similarity   100.0%;  Pred. No. 38;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      13 RAEGTEA 19
           |||||||
Db     537 RAEGTEA 543


RESULT  13
US-09-439-313-519
; Sequence 519, Application US/09439313
; Patent No. 6329505
; GENERAL INFORMATION:
;  APPLICANT: Xu, Jiangchun
;  APPLICANT:  Dillon, Davin C.
;  APPLICANT:  Mitcham, Jennifer L.
;  APPLICANT:  Harlocker, Susan Louise
;  APPLICANT:  Jiang Yuqui
;  APPLICANT:  Reed, Steven G.
;  APPLICANT:  Kalos, Michael
;  APPLICANT:  Fanger, Gary
;  APPLICANT:  Retter, Mark
;  APPLICANT:  Solk, John
```

; APPLICANT: Day, Craig
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
; FILE REFERENCE: 210121.427C9
; CURRENT APPLICATION NUMBER: US/09/439,313
; CURRENT FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 575
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 519
; LENGTH: 17
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Made in a lab
US-09-439-313-519

Query Match 21.4%; Score 6; DB 4; Length 17;
Best Local Similarity 100.0%; Pred. No. 6.9;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 16 GTEARR 21
||||||
Db 5 GTEARR 10

RESULT 14
US-09-290-640-66
; Sequence 66, Application US/09290640
; Patent No. 6204055
; GENERAL INFORMATION:
; APPLICANT: Dean, Nicholas M.
; APPLICANT: Marcusson, Eric G.
; TITLE OF INVENTION: Antisense Compound Modulation of Fas Mediated Signaling
; FILE REFERENCE: ISPH-0351
; CURRENT APPLICATION NUMBER: US/09/290,640
; CURRENT FILING DATE: 1999-04-12
; NUMBER OF SEQ ID NOS: 85
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 66
; LENGTH: 327
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-290-640-66

Query Match 21.4%; Score 6; DB 4; Length 327;
Best Local Similarity 100.0%; Pred. No. 79;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 EGLYQG 10
||||||
Db 46 EGLYQG 51

RESULT 15
5183745-4
;Patent No. 5183745
; APPLICANT: DANCHIN, ANTOINE;GLASER, PHILLIPPE;KRIN, EVELYN;
;BARZU, OCTAVIEN;LADANT, DANIEL;ULLMAN, AGNES
; TITLE OF INVENTION: ADENYL CYCLASE DERIVATIVES AND THEIR
;BIOLOGICAL USES
; NUMBER OF SEQUENCES: 13
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/426,541
; FILING DATE: 25-OCT-1989
;SEQ ID NO:4:
; LENGTH: 525
5183745-4

Query Match 21.4%; Score 6; DB 6; Length 525;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 16 GTEARR 21
||||||
Db 247 GTEARR 252

Search completed: June 26, 2002, 13:49:28
Job time: 306 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: June 26, 2002, 13:45:47 ; Search time 20.39 Seconds
(without alignments)
131.952 Million cell updates/sec

Title: US-09-838-785-25
Perfect score: 28
Sequence: 1 DFVGEGLYQGVPRAEGTEARRHYDEGVR 28

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 283138 seqs, 96089334 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : PIR_71:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	7	25.0	562	2	H69545	probable fatty-aci
2	7	25.0	569	2	C69471	probable fatty-aci
3	7	25.0	588	2	A75282	ABC transporter, A
4	7	25.0	1473	2	T31422	C-terminal domain-
5	7	25.0	1733	1	B45344	probable nuclear a
6	7	25.0	1958	2	B40505	hypothetical prote
7	6	21.4	77	2	B82467	hypothetical prote
8	6	21.4	136	2	A87681	conserved hypothet
9	6	21.4	219	2	F75402	probable competenc
10	6	21.4	231	2	G87039	conserved hypothet
11	6	21.4	252	1	A48008	electron transfer
12	6	21.4	264	2	D97625	hypothetical prote
13	6	21.4	264	2	AF2848	conserved hypothet
14	6	21.4	275	2	H87538	hypothetical prote
15	6	21.4	293	2	D89867	hypothetical prote
16	6	21.4	293	2	A71946	hypothetical prote
17	6	21.4	298	2	C64563	conserved hypothet
18	6	21.4	304	2	H83636	hypothetical prote
19	6	21.4	306	2	T26126	hypothetical prote
20	6	21.4	309	2	AB1346	hypothetical prote
21	6	21.4	309	2	AB1534	oxidoreductases ho
22	6	21.4	309	2	AE1716	hypothetical prote
23	6	21.4	309	2	AF1176	oxidoreductases ho
24	6	21.4	316	1	D71246	hypothetical prote
25	6	21.4	316	2	G72077	ct007 hypothetical
26	6	21.4	316	2	G86545	CT007 hypothetical
27	6	21.4	318	2	D75210	hypothetical prote
28	6	21.4	327	2	A46484	apoptosis-mediatin
29	6	21.4	348	2	T03530	cobW protein - Rho
30	6	21.4	352	2	F64751	probable ABC-type
31	6	21.4	355	2	AB3442	response regulator
32	6	21.4	358	2	E95842	probable lactose t
33	6	21.4	395	2	AE2143	hypothetical prote
34	6	21.4	418	2	AD3417	transporter, mfs s
35	6	21.4	420	2	AH1632	isocitrate dehyrog
36	6	21.4	458	2	F75367	potassium uptake p
37	6	21.4	471	2	T26612	hypothetical prote
38	6	21.4	483	2	A87583	peptidoglycan bind
39	6	21.4	488	2	F86916	NADH-dependent glu
40	6	21.4	508	2	A84854	hypothetical prote
41	6	21.4	577	2	H69354	probable fatty-aci
42	6	21.4	653	2	D87602	sensory box histid
43	6	21.4	667	2	AH2165	bicarbonate transp
44	6	21.4	686	2	A34612	zinc finger protei
45	6	21.4	687	2	B70515	hypothetical prote

ALIGNMENTS

RESULT 1
H69545
probable fatty-acid--CoA ligase (EC 6.2.1.-) fadD9 - Archaeoglobus fulgidus
C;Species: Archaeoglobus fulgidus
C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 21-Jul-2000
C;Accession: H69545
R;Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod
.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E
Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
Nature 390, 364-370, 1997
A;Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Artiach, P.; Kaine, B.P.; Sykes,
Smith, H.O.; Woese, C.R.; Venter, J.C.
A;Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing arch
A;Reference number: A69250; MUID:98049343
A;Accession: H69545
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-562 <KLE>
A;Cross-references: GB:AE001112; GB:AE000782; NID:g2689435; PIDN:AAB91290.1; PID:g265
C;Superfamily: 4-coumarate--CoA ligase; acetate--CoA ligase homology
C;Keywords: acid-thiol ligase
F;87-553/Domain: acetate--CoA ligase homology <ACL>

Query Match 25.0%; Score 7; DB 2; Length 562;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 22 HYDEGVR 28
|||||||
Db 27 HYDEGVR 33

RESULT 2
C69471
probable fatty-acid--CoA ligase (EC 6.2.1.-) fadD7 - Archaeoglobus fulgidus
C;Species: Archaeoglobus fulgidus
C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 21-Jul-2000
C;Accession: C69471
R;Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod
.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E
Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
Nature 390, 364-370, 1997
A;Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Artiach, P.; Kaine, B.P.; Sykes,
Smith, H.O.; Woese, C.R.; Venter, J.C.
A;Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing arch
A;Reference number: A69250; MUID:98049343
A;Accession: C69471
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-569 <KLE>
A;Cross-references: GB:AE000980; GB:AE000782; NID:g2689303; PIDN:AAB89478.1; PID:g264

C;Superfamily: 4-coumarate--CoA ligase; acetate--CoA ligase homology
C;Keywords: acid-thiol ligase
F;84-559/Domain: acetate--CoA ligase homology <ACL>

Query Match 25.0%; Score 7; DB 2; Length 569;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 22 HYDEGVR 28
|||||||
Db 24 HYDEGVR 30

RESULT 3
A75282
ABC transporter, ATP-binding protein, MsbA family - Deinococcus radiodurans (strain R1)
C;Species: Deinococcus radiodurans
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 26-May-2000
C;Accession: A75282
R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
, M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A;Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A;Reference number: A75250; MUID:20036896
A;Accession: A75282
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-588 <WHI>
A;Cross-references: GB:AE002068; GB:AE000513; NID:g6460186; PIDN:AAF11923.1; PID:g646019
A;Experimental source: strain R1
C;Genetics:
A;Gene: DR2379
A;Map position: 1
C;Superfamily: Escherichia coli ABC transporter mdlA; ATP-binding cassette homology

Query Match 25.0%; Score 7; DB 2; Length 588;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 15 EGTEARR 21
|||||||
Db 229 EGTEARR 235

RESULT 4
T31422
C-terminal domain-binding protein rA9 - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 07-Dec-1999
C;Accession: T31422
R;Yuryev, A.; Patturajan, M.; Litingtung, Y.; Joshi, R.V.; Gentile, C.; Gebara, M.; Corc
Proc. Natl. Acad. Sci. U.S.A. 93, 6975-6980, 1996
A;Title: The C-terminal domain of the largest subunit of RNA polymerase II interacts wit
A;Reference number: Z21024; MUID:96293459
A;Accession: T31422
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1473 <YUR>
A;Cross-references: EMBL:U49057; NID:g1438533; PID:g1438534; PIDN:AAC52658.1
A;Experimental source: hippocampus

Query Match 25.0%; Score 7; DB 2; Length 1473;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10 GVPRAEG 16
|||||||
Db 1276 GVPRAEG 1282

RESULT 5
B45344
probable nuclear antigen - suid herpesvirus 1 (strain Kaplan)
C;Species: suid herpesvirus 1
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 16-Jul-1999
C;Accession: B45344
R;Vlcek, C.; Kozmik, Z.; Paces, V.; Schirm, S.; Schwyzer, M.
Virology 179, 365-377, 1990
A;Title: Pseudorabies virus immediate-early gene overlaps with an oppositely oriented
A;Reference number: A45344; MUID:91021039
A;Accession: B45344
A;Status: translation not shown
A;Molecule type: DNA
A;Residues: 1-1733 <VLC>
A;Cross-references: GB:M34651; NID:g334070; PIDN:AAA47471.1; PID:g334072
C;Superfamily: pseudorabies virus 1 nuclear antigen

Query Match 25.0%; Score 7; DB 1; Length 1733;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 13 RAEGTEA 19
|||||||
Db 326 RAEGTEA 332

RESULT 6
B40505
hypothetical protein - suid herpesvirus 1 (strain Indiana-Funkhuser or Becker)
C;Species: suid herpesvirus 1
C;Date: 10-Apr-1992 #sequence_revision 10-Apr-1992 #text_change 02-Sep-2000
C;Accession: B40505
R;Cheung, A.K.
J. Virol. 65, 5260-5271, 1991
A;Title: Cloning of the latency gene and the early protein 0 gene of pseudorabies vir
A;Reference number: A40505; MUID:91374576
A;Accession: B40505
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1958 <CHE>
A;Cross-references: GB:M57505; NID:g334066; PIDN:AAA47468.1; PID:g334068
C;Superfamily: pseudorabies virus 1 nuclear antigen

Query Match 25.0%; Score 7; DB 2; Length 1958;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 13 RAEGTEA 19
|||||||
Db 537 RAEGTEA 543

RESULT 7
B82467
hypothetical protein VCA0389 [imported] - Vibrio cholerae (strain N16961 serogroup O1
C;Species: Vibrio cholerae
C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C;Accession: B82467
R;Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.
chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers
l, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A;Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A;Reference number: A82035; MUID:20406833
A;Accession: B82467
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-77 <HEI>
A;Cross-references: GB:AE004374; GB:AE003853; NID:g9657775; PIDN:AAF96295.1; GSPDB:GN
A;Experimental source: serogroup O1; strain N16961; biotype El Tor

```
C;Genetics:
A;Gene: VCA0389
A;Map position: 2


  Query Match             21.4%;  Score 6;  DB 2;  Length 77;
  Best Local Similarity   100.0%;  Pred. No. 21;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      2 FVGEGL 7
          ||||||
Db     65 FVGEGL 70


RESULT   8
A87681
conserved hypothetical protein CC3483 [imported] - Caulobacter crescentus
C;Species: Caulobacter crescentus
C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C;Accession: A87681
R;Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
 B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A;Title: Complete Genome Sequence of Caulobacter crescentus.
A;Reference number: A87249; MUID:21173698; PMID:11259647
A;Accession: A87681
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-136 <STO>
A;Cross-references: GB:AE005673; NID:g13425207; PIDN:AAK25445.1; GSPDB:GN00148
C;Genetics:
A;Gene: CC3483


  Query Match             21.4%;  Score 6;  DB 2;  Length 136;
  Best Local Similarity   100.0%;  Pred. No. 34;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      6 GLYQGV 11
          ||||||
Db     65 GLYQGV 70


RESULT   9
F75402
probable competence protein ComF - Deinococcus radiodurans (strain R1)
C;Species: Deinococcus radiodurans
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
C;Accession: F75402
R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
, M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A;Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A;Reference number: A75250; MUID:20036896
A;Accession: F75402
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-219 <WHI>
A;Cross-references: GB:AE001984; GB:AE000513; NID:g6459135; PIDN:AAF10957.1; PID:g645914
A;Experimental source: strain R1
C;Genetics:
A;Gene: DR1389
A;Map position: 1
C;Superfamily: transformation competence-related protein comF


  Query Match             21.4%;  Score 6;  DB 2;  Length 219;
  Best Local Similarity   100.0%;  Pred. No. 52;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      8 YQGVPR 13
          ||||||
Db     71 YQGVPR 76


RESULT  10
G87039
conserved hypothetical protein ML1045 [imported] - Mycobacterium leprae
C;Species: Mycobacterium leprae
C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C;Accession: G87039
R;Cole, S.T.; Eiglmeier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.;
R.; Davies, R.M.; Devlin, K.; Duthoy, S.; Feltwell, T.; Fraser, A.; Hamlin, N.; Holro
eam, M.A.; Rutherford, K.M.
Nature 409, 1007-1011, 2001
A;Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.;
A;Title: Massive gene decay in the leprosy bacillus.
A;Reference number: A86909; MUID:21128732; PMID:11234002
A;Accession: G87039
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-231 <STO>
A;Cross-references: GB:AL450380; NID:g13093058; PIDN:CAC31426.1; GSPDB:GN00147
C;Genetics:
A;Gene: ML1045


  Query Match             21.4%;  Score 6;  DB 2;  Length 231;
  Best Local Similarity   100.0%;  Pred. No. 54;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     17 TEARRH 22
          ||||||
Db    203 TEARRH 208


RESULT  11
A48008
electron transfer flavoprotein beta chain - Paracoccus denitrificans
C;Species: Paracoccus denitrificans
C;Date: 02-Jun-1995 #sequence_revision 12-Jul-1996 #text_change 08-May-1998
C;Accession: A48008; S23293
R;Bedzyk, L.A.; Escudero, K.W.; Gill, R.E.; Griffin, K.J.; Frerman, F.E.
J. Biol. Chem. 268, 20211-20217, 1993
A;Title: Cloning, sequencing, and expression of the genes encoding subunits of Paraco
A;Reference number: A48008; MUID:93388590
A;Accession: A48008
A;Molecule type: DNA
A;Residues: 1-252 <BED>
A;Cross-references: GB:L14864
R;Watmough, N.J.; Kiss, J.; Frerman, F.E.
Eur. J. Biochem. 205, 1089-1097, 1992
A;Title: Structural and redox relationships between Paracoccus denitrificans, porcine
A;Reference number: S23165; MUID:92249313
A;Accession: S23293
A;Molecule type: protein
A;Residues: 1-22,'R';33-49,'AVA',50-51;74-82;'Y',164-167,'E',169-183;189-197 <WAT>
C;Complex: heterodimer of alpha (see PIR:B48008) and beta chains that binds one molec
C;Function:
A;Description: electron transfer from any of several dehydrogenases to the respirator
C;Superfamily: electron transfer flavoprotein beta chain
C;Keywords: electron transfer; FAD; flavoprotein; heterodimer
F;1-252/Product: electron transfer flavoprotein beta chain #status predicted <MAT>


  Query Match             21.4%;  Score 6;  DB 1;  Length 252;
  Best Local Similarity   100.0%;  Pred. No. 58;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     13 RAEGTE 18
          ||||||
Db    110 RAEGTE 115
```

RESULT 12
D97625
hypothetical protein AGR_C_4025 [imported] - Agrobacterium tumefaciens (strain C58, Cere
C;Species: Agrobacterium tumefaciens
C;Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 11-Jan-2002
C;Accession: D97625
R;Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A;Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A;Reference number: A97359; PMID:11743194
A;Accession: D97625
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-264 <KUR>
A;Cross-references: GB:AE007869; PIDN:AAK87957.1; PID:g15157363; GSPDB:GN00169
C;Genetics:
A;Gene: AGR_C_4025
A;Map position: circular chromosome
C;Superfamily: Streptomyces coelicolor hypothetical protein SC4A10.14c

Query Match 21.4%; Score 6; DB 2; Length 264;
Best Local Similarity 100.0%; Pred. No. 61;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 EGLYQG 10
||||||
Db 36 EGLYQG 41

RESULT 13
AF2848
conserved hypothetical protein Atu2215 [imported] - Agrobacterium tumefaciens (strain C5
C;Species: Agrobacterium tumefaciens
C;Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 01-Feb-2002
C;Accession: AF2848
R;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I
erage, G.; Gillet, W.; Grant, C.; Guenthner, D.; Kutyavin, T.; Levy, R.; Li, M.; McClell
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A;Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A;Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A;Reference number: AB2577; PMID:11743193
A;Accession: AF2848
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-264 <KUR>
A;Cross-references: GB:AE008688; PIDN:AAL43204.1; PID:g17740685; GSPDB:GN00186
A;Experimental source: strain C58 (Dupont)
C;Genetics:
A;Gene: Atu2215
A;Map position: circular chromosome
C;Superfamily: Streptomyces coelicolor hypothetical protein SC4A10.14c

Query Match 21.4%; Score 6; DB 2; Length 264;
Best Local Similarity 100.0%; Pred. No. 61;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 EGLYQG 10
||||||
Db 36 EGLYQG 41

RESULT 14
H87538
hypothetical protein CC2337 [imported] - Caulobacter crescentus
C;Species: Caulobacter crescentus
C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 10-May-2001
C;Accession: H87538
R;Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Ko
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A;Title: Complete Genome Sequence of Caulobacter crescentus.
A;Reference number: A87249; MUID:21173698; PMID:11259647
A;Accession: H87538
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-275 <STO>
A;Cross-references: GB:AE005673; NID:g13423862; PIDN:AAK24308.1; GSPDB:GN00148
C;Genetics:
A;Gene: CC2337
C;Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology

Query Match 21.4%; Score 6; DB 2; Length 275;
Best Local Similarity 100.0%; Pred. No. 63;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9 QGVPRA 14
||||||
Db 153 QGVPRA 158

RESULT 15
D89867
hypothetical protein SA0854 [imported] - Staphylococcus aureus (strain N315)
C;Species: Staphylococcus aureus
C;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
C;Accession: D89867
R;Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; O
ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
Lancet 357, 1225-1240, 2001
A;Title: Whole genome sequencing of meticillin-resistant Stapylococcus aureus.
A;Reference number: A89758; MUID:21311952; PMID:11418146
A;Accession: D89867
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-293 <KUR>
A;Cross-references: GB:BA000018; PID:g13700799; PIDN:BAB42095.1; GSPDB:GN00149
A;Experimental source: strain N315
C;Genetics:
A;Gene: SA0854
C;Superfamily: oligopeptide permease protein oppB

Query Match 21.4%; Score 6; DB 2; Length 293;
Best Local Similarity 100.0%; Pred. No. 66;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 FVGEGL 7
||||||
Db 278 FVGEGL 283

Search completed: June 26, 2002, 13:50:03
Job time: 256 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: June 26, 2002, 13:48:22 ; Search time 10.1 Seconds
(without alignments)
107.341 Million cell updates/sec

Title: US-09-838-785-25
Perfect score: 28
Sequence: 1 DFVGEGLYQGVPRAEGTEARRHYDEGVR 28

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 105224 seqs, 38719550 residues

Word size : 0

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	7	25.0	1733	1	VNUA_PRVKA	P33485 pseudorabie
2	6	21.4	252	1	ETFB_PARDE	P38975 paracoccus
3	6	21.4	293	1	Y347_HELPJ	Q9zma1 helicobacte
4	6	21.4	298	1	Y347_HELPY	O25114 helicobacte
5	6	21.4	315	1	MFT_HUMAN	Q9h2d1 homo sapien
6	6	21.4	316	1	Y441_CHLPN	Q9z8a2 chlamydia p
7	6	21.4	327	1	TNR6_MOUSE	P25446 mus musculu
8	6	21.4	348	1	AFUC_ECOLI	P37009 escherichia
9	6	21.4	393	1	ACK2_RHIME	Q9x449 rhizobium m
10	6	21.4	686	1	ZN07_HUMAN	P17097 homo sapien
11	6	21.4	840	1	VPH1_YEAST	P32563 saccharomyc
12	6	21.4	959	1	SYL_AERPE	Q9yd97 aeropyrum p
13	6	21.4	1021	1	CARA_MOUSE	P58660 mus musculu
14	6	21.4	1363	1	VGR3_MOUSE	P35917 mus musculu
15	6	21.4	1705	1	CYAA_BORBR	Q57506 bordetella
16	6	21.4	1706	1	CYAA_BORPE	P15318 bordetella
17	5	17.9	60	1	HMEB_MYXGL	P31536 myxine glut
18	5	17.9	64	1	RL37_LYCES	P49212 lycopersico
19	5	17.9	66	1	ATPL_STRPN	Q59953 streptococc
20	5	17.9	70	1	CX2X_CONBE	Q9u3z3 conus betul
21	5	17.9	71	1	RECA_STRGC	P49986 streptococc
22	5	17.9	72	1	RL15_BACLI	P35138 bacillus li
23	5	17.9	85	1	COXG_BOVIN	P00429 bos taurus
24	5	17.9	85	1	COXG_HUMAN	P14854 homo sapien
25	5	17.9	90	1	PTSO_ECOLI	P33996 escherichia
26	5	17.9	92	1	DBH_CAUCR	O87475 caulobacter
27	5	17.9	94	1	YMFJ_ECOLI	P75973 escherichia
28	5	17.9	102	1	RECA_ENTFA	P42444 enterococcu
29	5	17.9	103	1	YH31_ARCFU	O28543 archaeoglob
30	5	17.9	104	1	RECA_LACDE	Q02347 lactobacill
31	5	17.9	104	1	RECA_LACHE	Q02348 lactobacill
32	5	17.9	104	1	RECA_LEUME	Q02349 leuconostoc
33	5	17.9	104	1	RECA_STRSL	P49987 streptococc
34	5	17.9	106	1	ATPR_DROME	Q24407 drosophila
35	5	17.9	116	1	Y998_RHIME	Q52966 rhizobium m
36	5	17.9	118	1	VAM1_HUMAN	P23763 homo sapien
37	5	17.9	118	1	VAM1_MOUSE	Q62442 mus musculu
38	5	17.9	118	1	VAM1_RAT	Q63666 rattus norv
39	5	17.9	118	1	VPX_SIVAI	Q02842 simian immu
40	5	17.9	118	1	Y357_VIBCH	Q9kv01 vibrio chol
41	5	17.9	120	1	CR13_ORYSA	Q06396 oryza sativ
42	5	17.9	121	1	RECA_RUMAL	P49985 ruminococcu
43	5	17.9	126	1	GP48_BPSP1	O48402 bacteriopha
44	5	17.9	126	1	YJ93_ARCFU	O28286 archaeoglob
45	5	17.9	131	1	INL3_CALJA	O97937 callithrix

ALIGNMENTS

RESULT 1
VNUA_PRVKA
ID VNUA_PRVKA STANDARD; PRT; 1733 AA.
AC P33485;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-FEB-1994 (Rel. 28, Last annotation update)
DE Probable nuclear antigen.
OS Pseudorabies virus (strain Kaplan) (PRV).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Varicellovirus.
OX NCBI_TaxID=33703;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91021039; PubMed=2171211;
RA Vlcek C., Kozmik Z., Paces V., Schirm S., Schwyzer M.;
RT "Pseudorabies virus immediate-early gene overlaps with an oppositely
RT oriented open reading frame: characterization of their promoter and
RT enhancer regions.";
RL Virology 179:365-377(1990).
CC ---
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC ---
DR EMBL; M34651; AAA47471.1; -.
DR PIR; B45344; B45344.
FT DOMAIN 112 117 POLY-THR.
FT DOMAIN 179 1733 GLY-RICH.
FT DOMAIN 192 196 POLY-SER.
FT DOMAIN 271 298 POLY-PRO.
FT DOMAIN 304 308 POLY-ARG.
FT DOMAIN 883 889 POLY-GLY.
FT DOMAIN 1398 1405 POLY-GLY.
SQ SEQUENCE 1733 AA; 172166 MW; 0C8CD8BE475BB5E2 CRC64;

Query Match 25.0%; Score 7; DB 1; Length 1733;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 13 RAEGTEA 19
|||||||
Db 326 RAEGTEA 332

RESULT 2
ETFB_PARDE
ID ETFB_PARDE STANDARD; PRT; 252 AA.
AC P38975;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)

```
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   Electron transfer flavoprotein beta-subunit (Beta-ETF) (Electron
DE   transfer flavoprotein small subunit) (ETFSS).
GN   ETFB.
OS   Paracoccus denitrificans.
OC   Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;
OC   Paracoccus.
OX   NCBI_TaxID=266;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 1-7.
RC   STRAIN=ATCC 13544;
RX   MEDLINE=93388590; PubMed=8376381;
RA   Bedzyk L.A., Escudero K.W., Gill R.E., Griffin K.J., Frerman F.E.;
RT   "Cloning, sequencing, and expression of the genes encoding subunits
RT   of Paracoccus denitrificans electron transfer flavoprotein.";
RL   J. Biol. Chem. 268:20211-20217(1993).
RN   [2]
RP   SEQUENCE OF 1-23; 33-51; 74-82; 163-183 AND 189-197.
RC   STRAIN=ATCC 13543;
RX   MEDLINE=92249313; PubMed=1576992;
RA   Watmough N.J., Kiss J., Frerman F.E.;
RT   "Structural and redox relationships between Paracoccus denitrificans,
RT   porcine and human electron-transferring flavoproteins.";
RL   Eur. J. Biochem. 205:1089-1097(1992).
RN   [3]
RP   X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS).
RX   MEDLINE=99151986; PubMed=10026281;
RA   Roberts D.L., Salazar D., Fulmer J.P., Frerman F.E., Kim J.-J.;
RT   "Crystal structure of Paracoccus denitrificans electron transfer
RT   flavoprotein: structural and electrostatic analysis of a conserved
RT   flavin binding domain.";
RL   Biochemistry 38:1977-1989(1999).
CC   -!- FUNCTION: THE ELECTRON TRANSFER FLAVOPROTEIN SERVES AS A SPECIFIC
CC       ELECTRON ACCEPTOR FOR OTHER DEHYDROGENASES. IT TRANSFERS THE
CC       ELECTRONS TO THE MAIN RESPIRATORY CHAIN VIA ETF-UBIQUINONE
CC       OXIDOREDUCTASE (ETF DEHYDROGENASE).
CC   -!- COFACTOR: CONTAINS ONE MOLECULE OF FAD PER DIMER.
CC   -!- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT.
CC   -!- SIMILARITY: BELONGS TO THE ETF BETA-SUBUNIT / FIXA FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L14864; AAA03071.1; -.
DR   PIR; S23293; S23293.
DR   PDB; 1EFP; 09-AUG-99.
DR   InterPro; IPR000049; ETF_beta.
DR   Pfam; PF01012; ETF_beta; 1.
DR   ProDom; PD003528; ETF_beta; 1.
DR   PROSITE; PS01065; ETF_BETA; 1.
KW   Electron transport; Flavoprotein; FAD; 3D-structure.
FT   CONFLICT     10     10       L -> V (IN REF. 2).
FT   CONFLICT     23     23       G -> R (IN REF. 2).
FT   CONFLICT    163    163       V -> Y (IN REF. 2).
FT   CONFLICT    168    168       Q -> E (IN REF. 2).
SQ   SEQUENCE   252 AA;  26673 MW;  00C43128BEA1EDED CRC64;


  Query Match             21.4%;  Score 6;  DB 1;  Length 252;
  Best Local Similarity   100.0%;  Pred. No. 23;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      13 RAEGTE 18
           ||||||
Db     110 RAEGTE 115


RESULT   3
Y347_HELPJ
ID   Y347_HELPJ     STANDARD;      PRT;   293 AA.
AC   Q9ZMA1;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical protein JHP0321.
GN   JHP0321.
OS   Helicobacter pylori J99 (Campylobacter pylori J99).
OC   Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC   Helicobacter.
OX   NCBI_TaxID=85963;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99120557; PubMed=9923682;
RA   Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C.,
RA   Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,
RA   Tummino P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,
RA   Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,
RA   Trust T.J.;
RT   "Genomic sequence comparison of two unrelated isolates of the human
RT   gastric pathogen Helicobacter pylori.";
RL   Nature 397:176-180(1999).
CC   -!- SIMILARITY: BELONGS TO THE RLU FAMILY OF PSEUDOURIDINE SYNTHASES.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE001468; AAD05911.1; -.
DR   InterPro; IPR002990; PSI_RLU.
DR   InterPro; IPR000613; PseudoU_synth.
DR   Pfam; PF00849; PseudoU_synth_2; 1.
DR   ProDom; PD001819; PseudoU_synth; 1.
DR   PROSITE; PS01129; PSI_RLU; FALSE_NEG.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   293 AA;  33668 MW;  D7EF3EC54A8E6168 CRC64;


  Query Match             21.4%;  Score 6;  DB 1;  Length 293;
  Best Local Similarity   100.0%;  Pred. No. 27;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 VGEGLY 8
             ||||||
Db     234 VGEGLY 239


RESULT   4
Y347_HELPY
ID   Y347_HELPY     STANDARD;      PRT;   298 AA.
AC   O25114;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical protein HP0347.
GN   HP0347.
OS   Helicobacter pylori (Campylobacter pylori).
OC   Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC   Helicobacter.
OX   NCBI_TaxID=210;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=26695 / ATCC 700392;
RX   MEDLINE=97394467; PubMed=9252185;
RA   Tomb J.-F., White O., Kerlavage A.R., Clayton R.A., Sutton G.G.,
RA   Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,
RA   Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
RA   Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,
```

```
RA   McKenney K., FitzGerald L.M., Lee N., Adams M.D., Hickey E.K.,
RA   Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
RA   Cotton M.D., Weidman J.M., Fujii C., Bowman C., Watthey L., Wallin E.,
RA   Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
RA   Venter J.C.;
RT   "The complete genome sequence of the gastric pathogen Helicobacter
RT   pylori.";
RL   Nature 388:539-547(1997).
CC   -!- SIMILARITY: BELONGS TO THE RLU FAMILY OF PSEUDOURIDINE SYNTHASES.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AE000551; AAD07404.1; -.
DR   TIGR; HP0347; -.
DR   InterPro; IPR002990; PSI_RLU.
DR   InterPro; IPR000613; PseudoU_synth.
DR   Pfam; PF00849; PseudoU_synth_2; 1.
DR   ProDom; PD001819; PseudoU_synth; 1.
DR   PROSITE; PS01129; PSI_RLU; FALSE_NEG.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   298 AA;  34347 MW;  116E4DD87779B15C CRC64;


  Query Match             21.4%;  Score 6;  DB 1;  Length 298;
  Best Local Similarity   100.0%;  Pred. No. 27;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 VGEGLY 8
           ||||||
Db     234 VGEGLY 239


RESULT   5
MFT_HUMAN
ID   MFT_HUMAN      STANDARD;      PRT;   315 AA.
AC   Q9H2D1;
DT   01-MAR-2002 (Rel. 41, Created)
DT   01-MAR-2002 (Rel. 41, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Mitochondrial folate transporter/carrier.
GN   MFT.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20538421; PubMed=10978331;
RA   Titus S.A., Moran R.G.;
RT   "Retrovirally mediated complementation of the glyB phenotype. Cloning
RT   of a human gene encoding the carrier for entry of folates into
RT   mitochondria.";
RL   J. Biol. Chem. 275:36811-36817(2000).
CC   -!- FUNCTION: Transport folate across the inner membranes of
CC       mitochondria.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein. Mitochondrial
CC       inner membrane.
CC   -!- DOMAIN: COMPOSED OF THREE HOMOLOGOUS DOMAINS.
CC   -!- TISSUE SPECIFICITY: Ubiquitous.
CC   -!- SIMILARITY: BELONGS TO THE MITOCHONDRIAL CARRIER FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AF283645; AAG37834.1; -.
DR   InterPro; IPR001993; Mitoch_carrier.
DR   InterPro; IPR002067; Mit_carrier.
DR   Pfam; PF00153; mito_carr; 3.
DR   PRINTS; PR00926; MITOCARRIER.
DR   PROSITE; PS00215; MITOCH_CARRIER; 1.
KW   Mitochondrion; Inner membrane; Repeat; Transmembrane; Transport.
FT   TRANSMEM     89    106       POTENTIAL.
FT   TRANSMEM    227    243       POTENTIAL.
FT   TRANSMEM    281    300       POTENTIAL.
SQ   SEQUENCE   315 AA;  35388 MW;  60EBC0D61951EB6E CRC64;


  Query Match             21.4%;  Score 6;  DB 1;  Length 315;
  Best Local Similarity   100.0%;  Pred. No. 28;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6 GLYQGV 11
           ||||||
Db      80 GLYQGV 85


RESULT   6
Y441_CHLPN
ID   Y441_CHLPN     STANDARD;      PRT;   316 AA.
AC   Q9Z8A2; Q9JRW4;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical protein CPn0441/CP0312/CPj0441.
GN   CPN0441 OR CP0312 OR CPJ0441.
OS   Chlamydia pneumoniae (Chlamydophila pneumoniae).
OC   Bacteria; Chlamydiales; Chlamydiaceae; Chlamydophila.
OX   NCBI_TaxID=83558;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CWL029;
RX   MEDLINE=99206606; PubMed=10192388;
RA   Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
RA   Olinger L., Grimwood J., Davis R.W., Stephens R.S.;
RT   "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";
RL   Nat. Genet. 21:385-389(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AR39;
RX   MEDLINE=20150255; PubMed=10684935;
RA   Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA   White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,
RA   Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA   Gwinn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
RA   Eisen J., Fraser C.M.;
RT   "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
RT   pneumoniae AR39.";
RL   Nucleic Acids Res. 28:1397-1406(2000).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=J138;
RX   MEDLINE=20330349; PubMed=10871362;
RA   Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
RA   Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
RT   "Comparison of whole genome sequences of Chlamydia pneumoniae J138
RT   from Japan and CWL029 from USA.";
RL   Nucleic Acids Res. 28:2311-2314(2000).
CC   -!- SIMILARITY: BELONGS TO THE CHLAMYDIAL CPN0441/CT007/TC0275
CC       FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
```

```
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AE001627; AAD18585.1; -.
DR   EMBL; AE002194; AAF73653.1; -.
DR   EMBL; AP002546; BAA98649.1; -.
DR   TIGR; CP0312; -.
KW   Hypothetical protein; Complete proteome.
FT   CONFLICT    243    243       Y -> H (IN REF. 1).
SQ   SEQUENCE   316 AA;  35415 MW;  C821CFB53B014B41 CRC64;


  Query Match             21.4%;  Score 6;  DB 1;  Length 316;
  Best Local Similarity   100.0%;  Pred. No. 28;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      6 GLYQGV 11
          ||||||
Db    160 GLYQGV 165


RESULT   7
TNR6_MOUSE
ID   TNR6_MOUSE     STANDARD;      PRT;   327 AA.
AC   P25446;
DT   01-MAY-1992 (Rel. 22, Created)
DT   01-MAY-1992 (Rel. 22, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Tumor necrosis factor receptor superfamily member 6 precursor (FASL
DE   receptor) (Apoptosis-mediating surface antigen FAS) (Apo-1 antigen)
DE   (CD95).
GN   TNFRSF6 OR APT1 OR FAS.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92148151; PubMed=1371136;
RA   Watanabe-Fukunaga R., Brannan C.I., Itoh N., Yonehara S.,
RA   Copeland N.G., Jenkins N.A., Nagata S.;
RT   "The cDNA structure, expression, and chromosomal assignment of the
RT   mouse Fas antigen.";
RL   J. Immunol. 148:1274-1279(1992).
RN   [2]
RP   SEQUENCE OF 1-96 FROM N.A.
RX   MEDLINE=93189576; PubMed=7680478;
RA   Adachi M., Watanabe-Fukunaga R., Nagata S.;
RT   "Aberrant transcription caused by the insertion of an early
RT   transposable element in an intron of the Fas antigen gene of lpr
RT   mice.";
RL   Proc. Natl. Acad. Sci. U.S.A. 90:1756-1760(1993).
RN   [3]
RP   VARIANT LPR.
RX   MEDLINE=92195401; PubMed=1372394;
RA   Watanabe-Fukunaga R., Brannan C.I., Copeland N.G., Jenkins N.A.,
RA   Nagata S.;
RT   "Lymphoproliferation disorder in mice explained by defects in Fas
RT   antigen that mediates apoptosis.";
RL   Nature 356:314-317(1992).
CC   -!- FUNCTION: RECEPTOR FOR A CYTOKINE LIGAND KNOWN AS FASL. THE
CC       ADAPTOR MOLECULE FADD RECRUITS CASPASE-8 TO THE ACTIVATED
CC       RECEPTOR. THE RESULTING AGGREGATE CALLED THE DEATH-INDUCING
CC       SIGNALING COMPLEX (DISC) PERFORMS CASPASE-8 PROTEOLYTIC
CC       ACTIVATION. ACTIVE CASPASE-8 INITIATES THE SUBSEQUENT CASCADE OF
CC       CASPASES (ASPARTATE-SPECIFIC CYSTEINE PROTEASES) MEDIATING
CC       APOPTOSIS. FAS-MEDIATED APOPTOSIS MAY HAVE A ROLE IN THE
CC       INDUCTION OF PERIPHERAL TOLERANCE, IN THE ANTIGEN-STIMULATED
CC       SUICIDE OF MATURE T-CELLS, OR BOTH (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- TISSUE SPECIFICITY: DETECTED IN VARIOUS TISSUES INCLUDING THYMUS,
CC       LIVER, LUNG, HEART, AND ADULT OVARY.
CC   -!- DOMAIN: CONTAINS A DEATH DOMAIN INVOLVED IN THE BINDING OF FADD,
CC       AND MAYBE TO OTHER CYTOSOLIC ADAPTOR PROTEINS.
CC   -!- DISEASE: DEFECTS IN TNFRSF6 ARE THE CAUSE OF A LYMPHOPROLIFERATION
CC       DISORDER (LPR) RESPONSIBLE FOR LYMPHADENOPATHY AND AUTOANTIBODY
CC       PRODUCTION.
CC   -!- SIMILARITY: CONTAINS 3 TNFR-CYS REPEATS.
CC   -!- SIMILARITY: CONTAINS 1 DEATH DOMAIN.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M83649; AAA37593.1; -.
DR   EMBL; S56490; AAB25700.1; -.
DR   EMBL; S56485; AAB25700.1; JOINED.
DR   EMBL; S56486; AAB25700.1; JOINED.
DR   PIR; A46484; A46484.
DR   HSSP; P25445; 1DDF.
DR   MGD; MGI:95484; Tnfrsf6.
DR   InterPro; IPR000488; Death.
DR   InterPro; IPR001368; TNFR_c6.
DR   Pfam; PF00531; death; 1.
DR   Pfam; PF00020; TNFR_c6; 3.
DR   SMART; SM00005; DEATH; 1.
DR   SMART; SM00208; TNFR; 3.
DR   PROSITE; PS00652; TNFR_NGFR_1; 2.
DR   PROSITE; PS50050; TNFR_NGFR_2; 2.
DR   PROSITE; PS50017; DEATH_DOMAIN; 1.
KW   Apoptosis; Receptor; Glycoprotein; Transmembrane; Repeat; Signal;
KW   Disease mutation.
FT   SIGNAL        1     21
FT   CHAIN        22    327       TUMOR NECROSIS FACTOR RECEPTOR
FT                                 SUPERFAMILY MEMBER 6.
FT   DOMAIN       22    169       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    170    186       POTENTIAL.
FT   DOMAIN      187    327       CYTOPLASMIC (POTENTIAL).
FT   REPEAT       43     79       TNFR-CYS 1.
FT   REPEAT       80    123       TNFR-CYS 2.
FT   REPEAT      124    162       TNFR-CYS 3.
FT   DOMAIN      222    306       DEATH.
FT   CARBOHYD     43     43       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    114    114       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARIANT     246    246       I -> N (IN LPR).
SQ   SEQUENCE   327 AA;  37418 MW;  F6BFFC5ACE356EEE CRC64;


  Query Match             21.4%;  Score 6;  DB 1;  Length 327;
  Best Local Similarity   100.0%;  Pred. No. 29;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      5 EGLYQG 10
          ||||||
Db     46 EGLYQG 51


RESULT   8
AFUC_ECOLI
ID   AFUC_ECOLI     STANDARD;      PRT;   348 AA.
AC   P37009; P77157;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Putative ferric transport ATP-binding protein afuC.
GN   AFUC OR B0262.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
```

```
RC   STRAIN=K12 / MG1655;
RX   MEDLINE=97426617; PubMed=9278503;
RA   Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA   Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA   Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA   Mau B., Shao Y.;
RT   "The complete genome sequence of Escherichia coli K-12.";
RL   Science 277:1453-1474(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / W3110;
RA   Takemoto K., Mori H., Murayama N., Kataoka K., Yano M., Itoh T.,
RA   Yamamoto Y., Inokuchi H., Miki T., Hatada E., Fukuda R., Ichihara S.,
RA   Mizuno T., Makino K., Nakata A., Yura T., Sampei G., Mizobuchi K.;
RT   "Systematic sequencing of the Escherichia coli genome: analysis of the
RT   4.0 - 6.0 min (189,987 - 281,416bp) region.";
RL   Submitted (NOV-1999) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Schramm S., Duncan M., Allen E., Araujo R., Aparicio A., Chung E.,
RA   Davis K., Federspiel N., Hyman R., Kalman S., Komp C., Kurdi O.,
RA   Lashkari D., Lew H., Lin D., Namath A., Oefner P., Roberts D.,
RA   Davis R.W.;
RL   Submitted (SEP-1996) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE OF 1-220 FROM N.A.
RX   MEDLINE=94156832; PubMed=8113168;
RA   Volkert M.R., Loewen P.C., Switala J., Crowley D., Conley M.;
RT   "The delta (argF-lacZ)205(U169) deletion greatly enhances resistance
RT   to hydrogen peroxide in stationary-phase Escherichia coli.";
RL   J. Bacteriol. 176:1297-1302(1994).
CC   -!- FUNCTION: PART OF THE BINDING-PROTEIN-DEPENDENT TRANSPORT SYSTEM
CC       FOR IRON. PROBABLY RESPONSIBLE FOR ENERGY COUPLING TO THE
CC       TRANSPORT SYSTEM.
CC   -!- MISCELLANEOUS: IN E.COLI, THE AFU SYSTEM SEEMS TO HAVE BEEN
CC       DELETED BY AN INSERTION SEQUENCE. AFUA IS TOTALLY LOST, AFUB IS
CC       PARTIALLY PRESENT AND AFUC IS TOTALLY CONSERVED.
CC   -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.
CC   -!- CAUTION: REF.4 DIFFERS FROM THAT SHOWN FROM POSITION 147 TO 220
CC       DUE TO A FRAMESHIFT.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE000134; AAC73365.1; ALT_INIT.
DR   EMBL; D83536; BAA77930.1; -.
DR   EMBL; U70214; AAB08683.1; ALT_INIT.
DR   EMBL; L20943; -; NOT_ANNOTATED_CDS.
DR   EcoGene; EG12340; afuC.
DR   InterPro; IPR003593; AAA.
DR   InterPro; IPR003439; ABC_transportr.
DR   InterPro; IPR001687; ATP_GTP_A.
DR   Pfam; PF00005; ABC_tran; 1.
DR   SMART; SM00382; AAA; 1.
DR   PROSITE; PS00211; ABC_TRANSPORTER; 1.
KW   Iron transport; Transport; ATP-binding; Complete proteome.
FT   NP_BIND      39     46       ATP (POTENTIAL).
FT   CONFLICT    147    148       LA -> WP (IN REF. 2).
SQ   SEQUENCE   348 AA;  39059 MW;  D4AD33754217DA1A CRC64;


  Query Match             21.4%;  Score 6;  DB 1;  Length 348;
  Best Local Similarity   100.0%;  Pred. No. 31;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 GVPRAE 15
           ||||||
Db     107 GVPRAE 112
```

```
RESULT   9
ACK2_RHIME
ID   ACK2_RHIME     STANDARD;      PRT;   393 AA.
AC   Q9X449; Q9XDG2;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Acetate kinase (EC 2.7.2.1) (Acetokinase).
GN   ACKA.
OS   Rhizobium meliloti (Sinorhizobium meliloti).
OC   Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC   Rhizobiaceae; Sinorhizobium.
OX   NCBI_TaxID=382;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=104A14;
RX   MEDLINE=99194732; PubMed=10094701;
RA   Summers M.L., Denton M.C., McDermott T.R.;
RT   "Genes coding for phosphotransacetylase and acetate kinase in
RT   Sinorhizobium meliloti are in an operon that is inducible by phosphate
RT   stress and controlled by phoB.";
RL   J. Bacteriol. 181:2217-2224(1999).
RN   [2]
RP   SEQUENCE OF 207-314 FROM N.A.
RC   STRAIN=104A14;
RX   MEDLINE=99022212; PubMed=9805396;
RA   Summers M.L., Elkins J.G., Elliot B.A., McDermott T.R.;
RT   "Expression and regulation of phosphate stress inducible genes in
RT   Sinorhizobium meliloti.";
RL   Mol. Plant Microbe Interact. 11:1094-1101(1998).
CC   -!- CATALYTIC ACTIVITY: ATP + acetate = ADP + acetyl phosphate.
CC   -!- PATHWAY: FIRST OF TWO STEPS IN THE CONVERSION OF ACETATE TO
CC       ACETYL-COA.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC   -!- SIMILARITY: BELONGS TO THE ACETOKINASE FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF095903; AAD24358.1; -.
DR   EMBL; AF074452; AAD42996.1; -.
DR   InterPro; IPR000890; Acetate_kin.
DR   Pfam; PF00871; Acetate_kinase; 1.
DR   PRINTS; PR00471; ACETATEKNASE.
DR   PROSITE; PS01075; ACETATE_KINASE_1; 1.
DR   PROSITE; PS01076; ACETATE_KINASE_2; FALSE_NEG.
KW   Transferase; Kinase.
FT   CONFLICT    209    212       SGAS -> AEFR (IN REF. 2).
FT   CONFLICT    227    238       LHRPSTGCRWDT -> GFTALDGLPMGTR (IN REF.
FT                                2).
SQ   SEQUENCE   393 AA;  42135 MW;  48FD185524CD6D4C CRC64;


  Query Match             21.4%;  Score 6;  DB 1;  Length 393;
  Best Local Similarity   100.0%;  Pred. No. 34;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      23 YDEGVR 28
           ||||||
Db     169 YDEGVR 174


RESULT  10
ZN07_HUMAN
ID   ZN07_HUMAN     STANDARD;      PRT;   686 AA.
AC   P17097; P17015;
```

```
DT   01-AUG-1990 (Rel. 15, Created)
DT   01-AUG-1990 (Rel. 15, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Zinc finger protein 7 (Zinc finger protein KOX4) (Zinc finger protein
DE   HF.16).
GN   ZNF7 OR KOX4.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Placenta;
RX   MEDLINE=90169993; PubMed=2106481;
RA   Lania L., Donti E., Pannuti A., Pascucci A., Pengue G.,
RA   Feliciello I., la Mantia G., Lanfrancone L., Pelicci P.-G.;
RT   "cDNA isolation, expression analysis, and chromosomal localization of
RT   two human zinc finger genes.";
RL   Genomics 6:333-340(1990).
RN   [2]
RP   SEQUENCE OF 413-468 FROM N.A.
RC   TISSUE=Lymphoid;
RX   MEDLINE=91145339; PubMed=2288909;
RA   Thiesen H.-J.;
RT   "Multiple genes encoding zinc finger domains are expressed in human T
RT   cells.";
RL   New Biol. 2:363-374(1990).
CC   -!- FUNCTION: MAY FUNCTION AS A TRANSCRIPTION FACTOR.
CC   -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC   -!- TISSUE SPECIFICITY: UBIQUITOUSLY PRESENT IN MANY HUMAN CELL LINES
CC       OF DIFFERENT EMBRYOLOGICAL DERIVATION.
CC   -!- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
CC       FINGER PROTEINS.
CC   -!- SIMILARITY: CONTAINS 1 KRAB DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M29580; AAA61313.1; -.
DR   EMBL; X52335; CAA36561.1; -.
DR   PIR; A34612; A34612.
DR   PIR; S10421; S10421.
DR   HSSP; P08047; 1SP2.
DR   MIM; 194531; -.
DR   InterPro; IPR001909; KRAB.
DR   InterPro; IPR000822; Znf-C2H2.
DR   Pfam; PF01352; KRAB; 1.
DR   Pfam; PF00096; zf-C2H2; 14.
DR   PRINTS; PR00048; ZINCFINGER.
DR   SMART; SM00349; KRAB; 1.
DR   SMART; SM00355; ZnF_C2H2; 14.
DR   PROSITE; PS50805; KRAB; 1.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2_1; 14.
DR   PROSITE; PS50157; ZINC_FINGER_C2H2_2; 14.
KW   Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;
KW   Nuclear protein; Repeat.
FT   DOMAIN       4     76       KRAB.
FT   DOMAIN     223    684       ZINC FINGERS.
FT   ZN_FING    223    245       C2H2-TYPE.
FT   ZN_FING    250    272       C2H2-TYPE.
FT   ZN_FING    278    300       C2H2-TYPE.
FT   ZN_FING    306    328       C2H2-TYPE.
FT   ZN_FING    334    356       C2H2-TYPE.
FT   ZN_FING    362    384       C2H2-TYPE.
FT   ZN_FING    413    435       C2H2-TYPE.
FT   ZN_FING    441    463       C2H2-TYPE.
FT   ZN_FING    469    491       C2H2-TYPE.
FT   ZN_FING    497    519       C2H2-TYPE.
FT   ZN_FING    525    547       C2H2-TYPE.
FT   ZN_FING    553    575       C2H2-TYPE.
FT   ZN_FING    581    603       C2H2-TYPE.
FT   ZN_FING    634    656       C2H2-TYPE.
FT   ZN_FING    662    684       C2H2-TYPE.
SQ   SEQUENCE   686 AA;  77887 MW;  0397ADDABEFD4FBF CRC64;


  Query Match             21.4%;  Score 6;  DB 1;  Length 686;
  Best Local Similarity   100.0%;  Pred. No. 56;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      14 AEGTEA 19
           ||||||
Db      71 AEGTEA 76


RESULT  11
VPH1_YEAST
ID   VPH1_YEAST     STANDARD;      PRT;   840 AA.
AC   P32563;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-OCT-1993 (Rel. 27, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   Vacuolar ATP synthase 95 kDa subunit (Vacuolar ATPase 95 kDa subunit).
GN   VPH1 OR YOR270C.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92332542; PubMed=1385813;
RA   Manolson M.F., Proteau D., Preston R.A., Stenbit A., Roberts B.T.,
RA   Hoyt M.A., Preuss D., Mulholland J., Botstein D., Jones E.W.;
RT   "The VPH1 gene encodes a 95-kDa integral membrane polypeptide
RT   required for in vivo assembly and activity of the yeast vacuolar
RT   H(+)-ATPase.";
RL   J. Biol. Chem. 267:14294-14303(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93147685; PubMed=1491220;
RA   Manolson M.F., Proteau D., Jones E.W.;
RT   "Evidence for a conserved 95-120 kDa subunit associated with and
RT   essential for activity of V-ATPases.";
RL   J. Exp. Biol. 172:105-112(1992).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C;
RX   MEDLINE=97051594; PubMed=8896271;
RA   Cheret G., Bernardi A., Sor F.J.;
RT   "DNA sequence analysis of the VPH1-SNF2 region on chromosome XV of
RT   Saccharomyces cerevisiae.";
RL   Yeast 12:1059-1064(1996).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C / FY1679;
RX   MEDLINE=97298311; PubMed=9153759;
RA   Poirey R., Jauniaux J.C.;
RT   "Sequencing analysis of a 36.8 kb fragment of yeast chromosome XV
RT   reveals 26 open reading frames including SEC63, CDC31, SUG2, GCD1,
RT   RBL2, PNT1, PAC1 and VPH1.";
RL   Yeast 13:483-487(1997).
CC   -!- FUNCTION: REQUIRED FOR ASSEMBLY AND ACTIVITY OF THE VACUOLAR
CC       ATPASE. POTENTIAL ROLE IN DIFFERENTIAL TARGETING AND REGULATION OF
CC       THE ENZYME FOR A SPECIFIC ORGANELLE.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein. Vacuolar.
CC   -!- SIMILARITY: BELONGS TO THE VATPASE 116 kDa SUBUNIT FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
```

```
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M89778; AAA35211.1; -.
DR   EMBL; X89633; CAA61776.1; -.
DR   EMBL; Z75178; CAA99494.1; -.
DR   EMBL; Z75179; CAA99496.1; -.
DR   PIR; A42970; A42970.
DR   SGD; S0005796; VPH1.
DR   InterPro; IPR002490; V_ATPase_sub_a.
DR   Pfam; PF01496; V_ATPase_sub_a; 1.
KW   Hydrogen ion transport; Transmembrane; Glycoprotein.
FT   DOMAIN        1    411       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    412    432       1 (POTENTIAL).
FT   DOMAIN      433    462       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    463    483       2 (POTENTIAL).
FT   DOMAIN      484    540       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    541    561       3 (POTENTIAL).
FT   DOMAIN      562    571       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    572    592       4 (POTENTIAL).
FT   DOMAIN      593    635       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    636    656       5 (POTENTIAL).
FT   DOMAIN      657    760       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    761    787       6 (POTENTIAL).
FT   DOMAIN      788    840       EXTRACELLULAR (POTENTIAL).
FT   CARBOHYD    113    113       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    280    280       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    324    324       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   840 AA;  95528 MW;  77709A914410CD4D CRC64;


  Query Match             21.4%;  Score 6;  DB 1;  Length 840;
  Best Local Similarity   100.0%;  Pred. No. 67;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      2 FVGEGL 7
          ||||||
Db    810 FVGEGL 815


RESULT  12
SYL_AERPE
ID   SYL_AERPE      STANDARD;      PRT;   959 AA.
AC   Q9YD97;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Leucyl-tRNA synthetase (EC 6.1.1.4) (Leucine--tRNA ligase) (LeuRS).
GN   LEUS OR APE1015.
OS   Aeropyrum pernix.
OC   Archaea; Crenarchaeota; Desulfurococcales; Desulfurococcaceae;
OC   Aeropyrum.
OX   NCBI_TaxID=56636;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K1;
RX   MEDLINE=99310339; PubMed=10382966;
RA   Kawarabayasi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,
RA   Jin-no K., Takahashi M., Sekine M., Baba S.-I., Ankai A., Kosugi H.,
RA   Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
RA   Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,
RA   Yamazaki J., Kushida N., Oguchi A., Aoki K.-I., Kubota K.,
RA   Nakamura Y., Nomura N., Sako Y., Kikuchi H.;
RT   "Complete genome sequence of an aerobic hyper-thermophilic
RT   crenarchaeon, Aeropyrum pernix K1.";
RL   DNA Res. 6:83-101(1999).
CC   -!- CATALYTIC ACTIVITY: ATP + L-leucine + tRNA(Leu) = AMP +
CC       diphosphate + L-leucyl-tRNA(Leu).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   -!- SIMILARITY: BELONGS TO CLASS-I AMINOACYL-TRNA SYNTHETASE FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AP000060; BAA80000.1; -.
DR   InterPro; IPR002300; tRNA-synt_1a.
DR   InterPro; IPR001412; tRNA-synt_I.
DR   Pfam; PF00133; tRNA-synt_1; 1.
DR   PROSITE; PS00178; AA_TRNA_LIGASE_I; FALSE_NEG.
KW   Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW   Complete proteome.
FT   SITE         39     49       "HIGH" REGION.
FT   SITE        637    641       "KMSKS" REGION.
FT   BINDING     640    640       ATP (BY SIMILARITY).
SQ   SEQUENCE   959 AA;  110519 MW;  985CB5168728F7AB CRC64;


  Query Match             21.4%;  Score 6;  DB 1;  Length 959;
  Best Local Similarity   100.0%;  Pred. No. 75;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      11 VPRAEG 16
           ||||||
Db    357 VPRAEG 362


RESULT  13
CARA_MOUSE
ID   CARA_MOUSE     STANDARD;      PRT;  1021 AA.
AC   P58660;
DT   01-MAR-2002 (Rel. 41, Created)
DT   01-MAR-2002 (Rel. 41, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Caspase recruitment domain protein 10 (Bcl10-interacting MAGUK protein
DE   1) (Bimp1).
GN   CARD10 OR BIMP1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21391892; PubMed=11387339;
RA   McAllister-Lucas L.M., Inohara N., Lucas P.C., Ruland J., Benito A.,
RA   Li Q., Chen S., Chen F.F., Yamaoka S., Verma I.M., Mak T.W.,
RA   Nunez G.;
RT   "Bimp1, a MAGUK family member linking protein kinase c activation to
RT   Bcl10-mediated NF-kappa B induction.";
RL   J. Biol. Chem. 276:30589-30597(2001).
CC   -!- FUNCTION: Activates NF-kappaB via Bcl10 and IKK.
CC   -!- SUBUNIT: CARD10 and Bcl10 bind to each other by CARD-CARD
CC       interaction. They both participate in a complex with MALT1, where
CC       MALT1 binds to Bcl10.
CC   -!- TISSUE SPECIFICITY: Highly expressed in kidney, heart followed by
CC       brain, lung, liver, skeletal muscle and testis.
CC   -!- SIMILARITY: CONTAINS 1 CARD DOMAIN.
CC   -!- CAUTION: Supposed to contain a SH3, a PDZ and a guanylate kinase-
CC       like domain. But none of these 3 domains are detected by PROSITE,
CC       Pfam or SMART.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AF363456; AAK60136.1; -.
DR   PROSITE; PS50209; CARD; 1.
```

```
KW   Coiled coil.
FT   DOMAIN       23    115       CARD.
FT   DOMAIN      138    450       COILED COIL (POTENTIAL).
FT   DOMAIN      558    565       POLY-SER.
SQ   SEQUENCE   1021 AA;  114413 MW;  4811A09BDB8F792C CRC64;


  Query Match             21.4%;  Score 6;  DB 1;  Length 1021;
  Best Local Similarity   100.0%;  Pred. No. 80;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      12 PRAEGT 17
           ||||||
Db     641 PRAEGT 646


RESULT  14
VGR3_MOUSE
ID   VGR3_MOUSE     STANDARD;      PRT;  1363 AA.
AC   P35917;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Vascular endothelial growth factor receptor 3 precursor (EC 2.7.1.112)
DE   (VEGFR-3) (Tyrosine-protein kinase receptor FLT4).
GN   FLT4 OR FLT-4.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J;
RX   MEDLINE=93330572; PubMed=8393164;
RA   Finnerty H., Kelleher K., Morris G.E., Bean K., Merberg D.M.,
RA   Kriz R., Morris J.C., Sookdeo H., Turner K.J., Wood C.R.;
RT   "Molecular cloning of murine FLT and FLT4.";
RL   Oncogene 8:2293-2298(1993).
CC   -!- FUNCTION: RECEPTOR FOR VEGF-C. HAS A TYROSINE-PROTEIN KINASE
CC       ACTIVITY.
CC   -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC       tyrosine phosphate.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- TISSUE SPECIFICITY: EXPRESSED IN ADULT LUNG AND LIVER, AND IN
CC       FETAL LIVER, BRAIN, INTESTINE AND PLACENTA.
CC   -!- SIMILARITY: BELONGS TO THE CSF-1/PDGF RECEPTOR FAMILY OF TYROSINE-
CC       PROTEIN KINASES.
CC   -!- SIMILARITY: CONTAINS 7 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; L07296; AAA40077.1; -.
DR   HSSP; P11362; 1FGK.
DR   MGD; MGI:95561; Flt4.
DR   InterPro; IPR000719; Euk_pkinase.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003600; Ig_like.
DR   InterPro; IPR001824; Receptor_tyr_kin_III.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF00047; ig; 6.
DR   Pfam; PF00069; pkinase; 2.
DR   SMART; SM00410; IG_like; 3.
DR   SMART; SM00408; IGc2; 2.
DR   SMART; SM00219; TyrKc; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR   PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR   PROSITE; PS00240; RECEPTOR_TYR_KIN_III; 1.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW   Transferase; Tyrosine-protein kinase; Phosphorylation; ATP-binding;
KW   Receptor; Transmembrane; Signal; Immunoglobulin domain; Repeat;
KW   Glycoprotein.
FT   SIGNAL        1     24       POTENTIAL.
FT   CHAIN        25   1363       VASCULAR ENDOTHELIAL GROWTH FACTOR
FT                                RECEPTOR 3.
FT   DOMAIN       25    775       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    776    797       POTENTIAL.
FT   DOMAIN      798   1363       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       44    118       IG-LIKE C2-TYPE DOMAIN 1.
FT   DOMAIN      151    213       IG-LIKE C2-TYPE DOMAIN 2.
FT   DOMAIN      245    317       IG-LIKE C2-TYPE DOMAIN 3.
FT   DOMAIN      351    403       IG-LIKE C2-TYPE DOMAIN 4.
FT   DOMAIN      438    541       IG-LIKE C2-TYPE DOMAIN 5.
FT   DOMAIN      571    660       IG-LIKE C2-TYPE DOMAIN 6.
FT   DOMAIN      692    758       IG-LIKE C2-TYPE DOMAIN 7.
FT   DOMAIN      845   1173       PROTEIN KINASE.
FT   NP_BIND     851    859       ATP (BY SIMILARITY).
FT   BINDING     879    879       ATP (BY SIMILARITY).
FT   ACT_SITE   1037   1037       BY SIMILARITY.
FT   DISULFID     51    111       POTENTIAL.
FT   DISULFID    158    206       POTENTIAL.
FT   DISULFID    252    310       POTENTIAL.
FT   DISULFID    445    534       POTENTIAL.
FT   DISULFID    578    653       POTENTIAL.
FT   DISULFID    699    751       POTENTIAL.
FT   MOD_RES    1068   1068       PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT   CARBOHYD     33     33       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    104    104       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    166    166       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    251    251       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    299    299       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    411    411       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    515    515       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    527    527       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    582    582       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    594    594       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    683    683       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    690    690       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    758    758       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1363 AA;  153015 MW;  F1BF8A2BDEF99BE9 CRC64;


  Query Match             21.4%;  Score 6;  DB 1;  Length 1363;
  Best Local Similarity   100.0%;  Pred. No. 1e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      15 EGTEAR 20
           ||||||
Db      85 EGTEAR 90


RESULT  15
CYAA_BORBR
ID   CYAA_BORBR     STANDARD;      PRT;  1705 AA.
AC   Q57506; O05179;
DT   01-NOV-1997 (Rel. 35, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Bifunctional hemolysin-adenylate cyclase precursor (Cyclolysin) (ACT)
DE   (AC-HLY) [Contains: Calmodulin-sensitive adenylate cyclase
DE   (EC 4.6.1.1) (ATP pyrophosphate-lyase) (Adenylyl cyclase); Hemolysin].
GN   CYA OR CYAA.
OS   Bordetella bronchiseptica (Alcaligenes bronchisepticus).
OC   Bacteria; Proteobacteria; beta subdivision; Alcaligenaceae;
OC   Bordetella.
OX   NCBI_TaxID=518;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CIP 9.73;
RX   MEDLINE=96009899; PubMed=7557410;
```

```
RA   Betsou F., Sismeiro O., Danchin A., Guiso N.;
RT   "Cloning and sequence of the Bordetella bronchiseptica adenylate
RT   cyclase-hemolysin-encoding gene: comparison with the Bordetella
RT   pertussis gene.";
RL   Gene 162:165-166(1995).
RN   [2]
RP   REVISIONS TO 1555-1558.
RA   Danchin A., Boursaux-Eude C.;
RL   Submitted (DEC-2000) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: THIS ADENYLATE CYCLASE BELONGS TO A SPECIAL CLASS OF
CC       BACTERIAL TOXIN. IT CAUSES WHOOPING COUGH BY ACTING ON MAMMALIAN
CC       CELLS BY ELEVATING CAMP-CONCENTRATION AND THUS DISRUPTS NORMAL
CC       CELL FUNCTION.
CC   -!- CATALYTIC ACTIVITY: ATP = 3',5'-cyclic AMP + diphosphate.
CC   -!- ENZYME REGULATION: ACTIVATED BY HOST CALMODULIN.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- DOMAIN: THE GLY-RICH REGION IS PROBABLY INVOLVED IN BINDING
CC       CALCIUM, WHICH IS REQUIRED FOR TARGET CELL-BINDING OR CYTOLYTIC
CC       ACTIVITY (BY SIMILARITY).
CC   -!- PTM: RELEASED IN A PROCESSED FORM.
CC   -!- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO ADENYLYL CYCLASE
CC       CLASS-2 FAMILY.
CC   -!- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE RTX
CC       PROKARYOTIC TOXIN FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z37112; CAA85481.2; -.
DR   InterPro; IPR001343; Hemlysn_Ca_bind.
DR   InterPro; IPR003355; RTX_N.
DR   InterPro; IPR003995; RtxA.
DR   Pfam; PF00353; hemolysinCabind; 5.
DR   Pfam; PF02382; RTX; 1.
DR   PRINTS; PR00313; CABNDNGRPT.
DR   PRINTS; PR01488; RTXTOXINA.
DR   PROSITE; PS00330; HEMOLYSIN_CALCIUM; 5.
KW   Lyase; cAMP synthesis; ATP-binding; Hemolysis; Toxin; Virulence;
KW   Whooping cough; Calcium-binding; Repeat; Lipoprotein; Palmitate.
FT   CHAIN         1    312       CALMODULIN-SENSITIVE ADENYLATE
FT                                CYCLASE.
FT   CHAIN       313   1705       HEMOLYSIN (BY SIMILARITY TO E.COLI
FT                                HEMOLYSIN HLYA).
FT   DOMAIN        1    399       A, CATALYTIC.
FT   DOMAIN      400    911       B, ALA/GLY-RICH.
FT   DOMAIN      912   1655       C.
FT   DOMAIN     1656   1705       D, ASP/GLY-RICH.
FT   NP_BIND     349    356       ATP (POTENTIAL).
FT   DOMAIN      912   1609       28 X 5 AA REPEATS, GLY-RICH.
FT   REPEAT      912    917       1.
FT   REPEAT     1014   1019       2.
FT   REPEAT     1023   1028       3.
FT   REPEAT     1032   1037       3.
FT   REPEAT     1041   1046       4.
FT   REPEAT     1050   1055       5.
FT   REPEAT     1059   1064       6.
FT   REPEAT     1079   1084       7.
FT   REPEAT     1164   1169       8.
FT   REPEAT     1173   1178       9.
FT   REPEAT     1182   1187       10.
FT   REPEAT     1202   1207       11.
FT   REPEAT     1279   1284       12.
FT   REPEAT     1288   1293       13.
FT   REPEAT     1297   1302       14.
FT   REPEAT     1306   1311       15.
FT   REPEAT     1315   1320       16.
FT   REPEAT     1324   1329       17.
FT   REPEAT     1344   1349       18.
FT   REPEAT     1420   1425       19.
FT   REPEAT     1429   1434       20.
FT   REPEAT     1438   1443       21.
FT   REPEAT     1447   1452       22.
FT   REPEAT     1555   1560       23.
FT   REPEAT     1564   1569       24.
FT   REPEAT     1573   1578       25.
FT   REPEAT     1582   1587       26.
FT   REPEAT     1592   1597       27.
FT   REPEAT     1604   1609       28.
FT   LIPID       859    859       PALMITATE (BY SIMILARITY).
FT   LIPID       982    982       PALMITATE (BY SIMILARITY).
SQ   SEQUENCE   1705 AA;  177153 MW;  C43B30F5886C835A CRC64;


  Query Match             21.4%;  Score 6;  DB 1;  Length 1705;
  Best Local Similarity   100.0%;  Pred. No. 1.2e+02;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      16 GTEARR 21
           ||||||
Db     254 GTEARR 259


Search completed: June 26, 2002, 13:51:04
Job time: 162 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: June 26, 2002, 13:47:32 ; Search time 26.53 Seconds
(without alignments)
182.580 Million cell updates/sec

Title: US-09-838-785-25
Perfect score: 28
Sequence: 1 DFVGEGLYQGVPRAEGTEARRHYDEGVR 28

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 562222 seqs, 172994929 residues

Word size : 0

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SPTREMBL_19:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	15	53.6	501	6	Q95KC5	Q95kc5 macaca fasc
2	15	53.6	553	4	Q96JT2	Q96jt2 homo sapien
3	15	53.6	553	6	Q95KI5	Q95ki5 macaca fasc
4	7	25.0	179	5	Q9VSG8	Q9vsg8 drosophila
5	7	25.0	364	2	Q9RJ80	Q9rj80 streptomyce
6	7	25.0	562	17	O30302	O30302 archaeoglob
7	7	25.0	569	17	O28502	O28502 archaeoglob
8	7	25.0	588	16	Q9RRV6	Q9rrv6 deinococcus
9	7	25.0	1473	11	Q63625	Q63625 rattus norv
10	7	25.0	1958	12	Q69340	Q69340 pseudorabie
11	6	21.4	77	16	Q9KMG7	Q9kmg7 vibrio chol
12	6	21.4	80	2	Q93H79	Q93h79 streptomyce
13	6	21.4	83	2	Q9ZIK4	Q9zik4 helicobacte
14	6	21.4	123	6	Q9GMX9	Q9gmx9 macaca fasc
15	6	21.4	133	11	Q9CS07	Q9cs07 mus musculu
16	6	21.4	136	16	Q9A2S3	Q9a2s3 caulobacter
17	6	21.4	149	6	Q9BGT5	Q9bgt5 macaca fasc
18	6	21.4	150	2	Q47303	Q47303 escherichia
19	6	21.4	150	2	Q93PW5	Q93pw5 streptococc
20	6	21.4	164	11	Q9CQ46	Q9cq46 mus musculu
21	6	21.4	167	2	Q9RKA3	Q9rka3 streptomyce
22	6	21.4	206	15	Q9Q080	Q9q080 chimpanzee
23	6	21.4	219	16	Q9RUJ7	Q9ruj7 deinococcus
24	6	21.4	231	16	Q9CCB2	Q9ccb2 mycobacteri
25	6	21.4	249	2	Q9S338	Q9s338 prochloroco
26	6	21.4	262	16	Q98G03	Q98g03 rhizobium l
27	6	21.4	266	2	Q9Z6G6	Q9z6g6 vibrio para
28	6	21.4	269	11	Q9D0B9	Q9d0b9 mus musculu
29	6	21.4	275	16	Q9A5W0	Q9a5w0 caulobacter
30	6	21.4	293	16	Q99V95	Q99v95 staphylococ
31	6	21.4	297	2	Q50009	Q50009 mycobacteri
32	6	21.4	300	16	Q98CI4	Q98ci4 rhizobium l
33	6	21.4	303	6	Q95KC6	Q95kc6 macaca fasc
34	6	21.4	304	16	Q9I762	Q9i762 pseudomonas
35	6	21.4	306	5	Q23151	Q23151 caenorhabdi
36	6	21.4	309	16	Q92DK3	Q92dk3 listeria in
37	6	21.4	309	16	Q929K2	Q929k2 listeria in
38	6	21.4	315	4	Q9H2D1	Q9h2d1 homo sapien
39	6	21.4	315	4	Q96SU7	Q96su7 homo sapien
40	6	21.4	315	4	Q96JZ6	Q96jz6 homo sapien
41	6	21.4	315	6	Q9BE05	Q9be05 macaca fasc
42	6	21.4	315	6	Q95J75	Q95j75 macaca fasc
43	6	21.4	316	17	O57966	O57966 pyrococcus
44	6	21.4	318	17	Q9V267	Q9v267 pyrococcus
45	6	21.4	327	11	Q9DCQ1	Q9dcq1 mus musculu

ALIGNMENTS

```
RESULT   1
Q95KC5
ID   Q95KC5      PRELIMINARY;      PRT;   501 AA.
AC   Q95KC5;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   HYPOTHETICAL 53.4 KDA PROTEIN.
OS   Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
OC   Cercopithecinae; Macaca.
OX   NCBI_TaxID=9541;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=MEDULLA OBLOGATA;
RA   Osada N., Hida M., Kusuda J., Tanuma R., Iseki K., Hirai M., Terao K.,
RA   Suzuki Y., Sugano S., Hashimoto K.;
RT   "Isolation of full-length cDNA clones from macaque brain cDNA
RT   libraries.";
RL   Submitted (JUN-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB062977; BAB60745.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   501 AA;  53447 MW;  8C554BBD04EE0470 CRC64;


  Query Match            53.6%;  Score 15;  DB 6;  Length 501;
  Best Local Similarity  100.0%;  Pred. No. 1.9e-07;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 DFVGEGLYQGVPRAE 15
           |||||||||||||||
Db     242 DFVGEGLYQGVPRAE 256


RESULT   2
Q96JT2
ID   Q96JT2      PRELIMINARY;      PRT;   553 AA.
AC   Q96JT2;
```

```
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   PROSTEIN.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=PROSTATE;
RX   MEDLINE=21139094; PubMed=11245466;
RA   Xu J., Kalos M., Stolk J.A., Zasloff E.J., Zhang X., Houghton R.L.,
RA   Filho A.M., Nolasco M., Badaro R., Reed S.G.;
RT   "Identification and characterization of prostein, a novel prostate-
RT   specific protein.";
RL   Cancer Res. 61:1563-1568(2001).
DR   EMBL; AY033593; AAK54386.1; -.
SQ   SEQUENCE   553 AA;  59322 MW;  0AFA23FBC742A667 CRC64;


  Query Match             53.6%;  Score 15;  DB 4;  Length 553;
  Best Local Similarity   100.0%;  Pred. No. 2.1e-07;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 DFVGEGLYQGVPRAE 15
          |||||||||||||||
Db    294 DFVGEGLYQGVPRAE 308


RESULT   3
Q95KI5
ID   Q95KI5      PRELIMINARY;      PRT;   553 AA.
AC   Q95KI5;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   HYPOTHETICAL 59.4 KDA PROTEIN.
OS   Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
OC   Cercopithecinae; Macaca.
OX   NCBI_TaxID=9541;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=TEMPORAL LOBE RIGHT;
RA   Osada N., Hida M., Kusuda J., Tanuma R., Iseki K., Hirai M., Terao K.,
RA   Suzuki Y., Sugano S., Hashimoto K.;
RT   "Isolation of full-length cDNA clones from macaque brain cDNA
RT   libraries.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB060851; BAB46871.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   553 AA;  59392 MW;  0718F3A91FB3BF1E CRC64;


  Query Match             53.6%;  Score 15;  DB 6;  Length 553;
  Best Local Similarity   100.0%;  Pred. No. 2.1e-07;
  Matches   15;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 DFVGEGLYQGVPRAE 15
          |||||||||||||||
Db    294 DFVGEGLYQGVPRAE 308


RESULT   4
Q9VSG8
ID   Q9VSG8      PRELIMINARY;      PRT;   179 AA.
AC   Q9VSG8;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   CG7197 PROTEIN.
GN   CG7197.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
DR   EMBL; AE003556; AAF50451.1; -.
DR   HSSP; P32889; 1HUR.
DR   FlyBase; FBgn0035866; CG7197.
DR   InterPro; IPR000251; ARF.
DR   InterPro; IPR001806; Ras_trnsfrmng.
DR   InterPro; IPR002046; Sar1_GTPBP.
DR   Pfam; PF00025; arf; 1.
DR   PRINTS; PR00449; RASTRNSFRMNG.
DR   PRINTS; PR00328; SAR1GTPBP.
DR   SMART; SM00177; ARF; 1.
DR   PROSITE; PS01019; ARF; 1.
KW   GTP-binding.
SQ   SEQUENCE   179 AA;  20626 MW;  74DDA1430BD5CDAC CRC64;


  Query Match             25.0%;  Score 7;  DB 5;  Length 179;
  Best Local Similarity   100.0%;  Pred. No. 11;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      4 GEGLYQG 10
          |||||||
Db    162 GEGLYQG 168


RESULT   5
```

```
Q9RJ80
ID   Q9RJ80      PRELIMINARY;      PRT;   364 AA.
AC   Q9RJ80;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   GLYCINE BETAINE TRANSPORT ATP-BINDING PROTEIN.
GN   OPUAA.
OS   Streptomyces coelicolor.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=1902;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RA   Seeger K.J., Harris D.;
RL   Submitted (OCT-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RA   Cerdeno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
RL   Submitted (OCT-1999) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RX   MEDLINE=97000351; PubMed=8843436;
RA   Redenbach M., Kieser H.M., Denapaite D., Eichner A., Cullum J.,
RA   Kinashi H., Hopwood D.A.;
RT   "A set of ordered cosmids and a detailed genetic and physical map for
RT   the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL   Mol. Microbiol. 21:77-96(1996).
CC   -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.
DR   EMBL; AL132648; CAB59474.1; -.
DR   InterPro; IPR003593; AAA.
DR   InterPro; IPR003439; ABC_transportr.
DR   InterPro; IPR001687; ATP_GTP_A.
DR   InterPro; IPR000644; CBS.
DR   Pfam; PF00005; ABC_tran; 1.
DR   SMART; SM00382; AAA; 1.
DR   SMART; SM00116; CBS; 1.
DR   PROSITE; PS00211; ABC_TRANSPORTER; 1.
KW   ATP-binding; Transport.
SQ   SEQUENCE   364 AA;  39970 MW;  B67DE74839610491 CRC64;


  Query Match             25.0%;  Score 7;  DB 2;  Length 364;
  Best Local Similarity   100.0%;  Pred. No. 21;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       9 QGVPRAE 15
           |||||||
Db     134 QGVPRAE 140


RESULT   6
O30302
ID   O30302      PRELIMINARY;      PRT;   562 AA.
AC   O30302;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   LONG-CHAIN-FATTY-ACID--COA LIGASE (FADD-9).
GN   AF2368.
OS   Archaeoglobus fulgidus.
OC   Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;
OC   Archaeoglobus.
OX   NCBI_TaxID=2234;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX   MEDLINE=98049343; PubMed=9389475;
RA   Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA   Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
RA   Richardson D.L., Kerlavage A.R., Graham D.E., Kyrpides N.C.,
RA   Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA   Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA   Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
RA   Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
RA   Cotton M.D., Spriggs T., Artiach P., Kaine B.P., Sykes S.M.,
RA   Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA   Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA   Venter J.C.;
RT   "The complete genome sequence of the hyperthermophilic, sulphate-
RT   reducing archaeon Archaeoglobus fulgidus.";
RL   Nature 390:364-370(1997).
DR   EMBL; AE001112; AAB91290.1; -.
DR   HSSP; P08659; 1LCI.
DR   TIGR; AF2368; -.
DR   InterPro; IPR000873; AMP-bind.
DR   Pfam; PF00501; AMP-binding; 1.
DR   PROSITE; PS00455; AMP_BINDING; 1.
KW   Hypothetical protein; Ligase; Complete proteome.
SQ   SEQUENCE   562 AA;  62487 MW;  CCBC8944EFF9DA2B CRC64;


  Query Match             25.0%;  Score 7;  DB 17;  Length 562;
  Best Local Similarity   100.0%;  Pred. No. 30;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      22 HYDEGVR 28
           |||||||
Db      27 HYDEGVR 33


RESULT   7
O28502
ID   O28502      PRELIMINARY;      PRT;   569 AA.
AC   O28502;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   LONG-CHAIN-FATTY-ACID--COA LIGASE (FADD-7).
GN   AF1772.
OS   Archaeoglobus fulgidus.
OC   Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;
OC   Archaeoglobus.
OX   NCBI_TaxID=2234;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX   MEDLINE=98049343; PubMed=9389475;
RA   Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA   Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
RA   Richardson D.L., Kerlavage A.R., Graham D.E., Kyrpides N.C.,
RA   Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA   Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA   Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
RA   Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
RA   Cotton M.D., Spriggs T., Artiach P., Kaine B.P., Sykes S.M.,
RA   Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA   Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA   Venter J.C.;
RT   "The complete genome sequence of the hyperthermophilic, sulphate-
RT   reducing archaeon Archaeoglobus fulgidus.";
RL   Nature 390:364-370(1997).
DR   EMBL; AE000980; AAB89478.1; -.
DR   HSSP; P08659; 1LCI.
DR   TIGR; AF1772; -.
DR   InterPro; IPR000873; AMP-bind.
DR   Pfam; PF00501; AMP-binding; 1.
DR   PROSITE; PS00455; AMP_BINDING; UNKNOWN_1.
KW   Hypothetical protein; Ligase; Complete proteome.
SQ   SEQUENCE   569 AA;  64316 MW;  467F5668A87DC478 CRC64;


  Query Match             25.0%;  Score 7;  DB 17;  Length 569;
```

```
  Best Local Similarity   100.0%;  Pred. No. 30;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      22 HYDEGVR 28
           |||||||
Db      24 HYDEGVR 30


RESULT   8
Q9RRV6
ID   Q9RRV6      PRELIMINARY;      PRT;   588 AA.
AC   Q9RRV6;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE   ABC TRANSPORTER, ATP-BINDING PROTEIN, MSBA FAMILY.
GN   DR2379.
OS   Deinococcus radiodurans.
OC   Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.
OX   NCBI_TaxID=1299;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=R1;
RX   MEDLINE=20036896; PubMed=10567266;
RA   White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA   Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA   Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA   Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,
RA   Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA   Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA   Fraser C.M.;
RT   "Genome sequence of the radioresistant bacterium Deinococcus
RT   radiodurans R1.";
RL   Science 286:1571-1577(1999).
CC   -!- SIMILARITY: BELONGS TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY
CC       (ABC TRANSPORTERS).
DR   EMBL; AE002068; AAF11923.1; -.
DR   TIGR; DR2379; -.
DR   InterPro; IPR003593; AAA.
DR   InterPro; IPR001140; ABC_transporter_tmem.
DR   InterPro; IPR003439; ABC_transportr.
DR   InterPro; IPR001360; Glyco_hydro_1.
DR   Pfam; PF00664; ABC_membrane; 1.
DR   Pfam; PF00005; ABC_tran; 1.
DR   SMART; SM00382; AAA; 1.
DR   PROSITE; PS00211; ABC_TRANSPORTER; 1.
DR   PROSITE; PS00572; GLYCOSYL_HYDROL_F1_1; UNKNOWN_1.
KW   ATP-binding; Complete proteome; Transport.
SQ   SEQUENCE   588 AA;  64714 MW;  DCB107824C9FBFF5 CRC64;


  Query Match             25.0%;  Score 7;  DB 16;  Length 588;
  Best Local Similarity   100.0%;  Pred. No. 31;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      15 EGTEARR 21
           |||||||
Db     229 EGTEARR 235


RESULT   9
Q63625
ID   Q63625      PRELIMINARY;      PRT;  1473 AA.
AC   Q63625;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   RA9.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=HIPPOCAMPUS;
RX   MEDLINE=96293459; PubMed=8692929;
RA   Yuryev A., Patturajan M., Litingtung Y., Joshi R.V., Gentile C.,
RA   Gebara M., Corden J.L.;
RT   "The C-terminal domain of the largest subunit of RNA polymerase II
RT   interacts with a novel set of serine/arginine-rich proteins.";
RL   Proc. Natl. Acad. Sci. U.S.A. 93:6975-6980(1996).
DR   EMBL; U49057; AAC52658.1; -.
SQ   SEQUENCE   1473 AA;  161204 MW;  949EE6F5873989BF CRC64;


  Query Match             25.0%;  Score 7;  DB 11;  Length 1473;
  Best Local Similarity   100.0%;  Pred. No. 66;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      10 GVPRAEG 16
           |||||||
Db    1276 GVPRAEG 1282


RESULT  10
Q69340
ID   Q69340      PRELIMINARY;      PRT;  1958 AA.
AC   Q69340;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   ORF1, ORF2, AND ORF3.
OS   Pseudorabies virus.
OC   Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC   Alphaherpesvirinae; Varicellovirus.
OX   NCBI_TaxID=10345;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=INDIANA-FUNKHAUSER;
RX   MEDLINE=91374576; PubMed=1654441;
RA   Cheung A.K.;
RT   "Cloning of the latency gene and the early protein 0 gene of
RT   pseudorabies virus.";
RL   J. Virol. 65:5260-5271(1991).
DR   EMBL; M57505; AAA47468.1; -.
SQ   SEQUENCE   1958 AA;  193474 MW;  0DEC36EDC29E2E68 CRC64;


  Query Match             25.0%;  Score 7;  DB 12;  Length 1958;
  Best Local Similarity   100.0%;  Pred. No. 84;
  Matches    7;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      13 RAEGTEA 19
           |||||||
Db     537 RAEGTEA 543


RESULT  11
Q9KMG7
ID   Q9KMG7      PRELIMINARY;      PRT;    77 AA.
AC   Q9KMG7;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   HYPOTHETICAL PROTEIN VCA0389.
GN   VCA0389.
OS   Vibrio cholerae.
OC   Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX   NCBI_TaxID=666;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=EL TOR N16961 / SEROTYPE O1;
RX   MEDLINE=20406833; PubMed=10952301;
RA   Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA   Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA   Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
```

```
RA   Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA   McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
RA   Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA   Fraser C.M.;
RT   "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT   cholerae.";
RL   Nature 406:477-483(2000).
DR   EMBL; AE004374; AAF96295.1; -.
DR   TIGR; VCA0389; -.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   77 AA;  9013 MW;  CF2172BF4B38D124 CRC64;


  Query Match             21.4%;  Score 6;  DB 16;  Length 77;
  Best Local Similarity   100.0%;  Pred. No. 59;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2 FVGEGL 7
           ||||||
Db      65 FVGEGL 70


RESULT  12
Q93H79
ID   Q93H79      PRELIMINARY;      PRT;    80 AA.
AC   Q93H79;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   FERREDOXINE.
GN   PTEE.
OS   Streptomyces avermitilis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=33903;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21477403; PubMed=11572948;
RA   Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
RA   Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osonoe T.,
RA   Kikuchi H., Shiba T., Sakaki Y., Hattori M.;
RT   "Genome sequence of an industrial microorganism Streptomyces
RT   avermitilis: Deducing the ability of producing secondary
RT   metabolites.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).
DR   EMBL; AB070949; BAB69311.1; -.
SQ   SEQUENCE   80 AA;  8475 MW;  B12A8B1B9D25A556 CRC64;


  Query Match             21.4%;  Score 6;  DB 2;  Length 80;
  Best Local Similarity   100.0%;  Pred. No. 60;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      13 RAEGTE 18
           ||||||
Db       4 RAEGTE 9


RESULT  13
Q9ZIK4
ID   Q9ZIK4      PRELIMINARY;      PRT;    83 AA.
AC   Q9ZIK4;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   HYPOTHETICAL 9.5 KDA PROTEIN (FRAGMENT).
OS   Helicobacter pylori (Campylobacter pylori).
OC   Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC   Helicobacter.
OX   NCBI_TaxID=210;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=J166;
RX   MEDLINE=99007275; PubMed=9789049;
RA   Akopyants N.S., Fradkov A., Diatchenko L., Hill J.E., Siebert P.D.,
RA   Lukyanov S.A., Sverdlov E.D., Berg D.E.;
RT   "PCR-based subtractive hybridization and differences in gene content
RT   among strains of Helicobacter pylori.";
RL   Proc. Natl. Acad. Sci. U.S.A. 95:13108-13113(1998).
DR   EMBL; AF025970; AAC69259.1; -.
DR   InterPro; IPR000613; PseudoU_synth.
DR   ProDom; PD001819; PseudoU_synth; 1.
KW   Hypothetical protein.
FT   NON_TER       1      1
SQ   SEQUENCE   83 AA;  9503 MW;  795E87B3ED183EFF CRC64;


  Query Match             21.4%;  Score 6;  DB 2;  Length 83;
  Best Local Similarity   100.0%;  Pred. No. 62;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3 VGEGLY 8
           ||||||
Db      24 VGEGLY 29


RESULT  14
Q9GMX9
ID   Q9GMX9      PRELIMINARY;      PRT;   123 AA.
AC   Q9GMX9;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   HYPOTHETICAL 13.6 KDA PROTEIN.
OS   Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
OC   Cercopithecinae; Macaca.
OX   NCBI_TaxID=9541;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN PARIETAL LOBE;
RA   Osada N., Hida M., Kusuda J., Tanuma R., Iseki K., Hirai M., Terao K.,
RA   Suzuki Y., Sugano S., Hashimoto K.;
RT   "Isolation of full-length cDNA clones from macaque brain cDNA
RT   libraries.";
RL   Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB047596; BAB12120.1; -.
DR   InterPro; IPR001993; Mitoch_carrier.
DR   Pfam; PF00153; mito_carr; 1.
DR   PROSITE; PS00215; MITOCH_CARRIER; UNKNOWN_1.
KW   Hypothetical protein.
SQ   SEQUENCE   123 AA;  13618 MW;  2B553EA863545098 CRC64;


  Query Match             21.4%;  Score 6;  DB 6;  Length 123;
  Best Local Similarity   100.0%;  Pred. No. 87;
  Matches    6;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       6 GLYQGV 11
           ||||||
Db      80 GLYQGV 85


RESULT  15
Q9CS07
ID   Q9CS07      PRELIMINARY;      PRT;   133 AA.
AC   Q9CS07;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   1700073K01RIK PROTEIN (FRAGMENT).
GN   1700073K01RIK.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
```

OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=EMBRYONIC LIVER;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AK010880; BAB27243.1; -.
DR HSSP; P02593; 1CDM.
DR MGD; MGI:1915476; 1700073K01Rik.
DR InterPro; IPR002048; EF-hand.
DR SMART; SM00054; EFh; 2.
FT NON_TER 133 133
SQ SEQUENCE 133 AA; 15322 MW; 7D23323D63F7F170 CRC64;

Query Match 21.4%; Score 6; DB 11; Length 133;
Best Local Similarity 100.0%; Pred. No. 92;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      14 AEGTEA 19
           ||||||
Db       7 AEGTEA 12
```

Search completed: June 26, 2002, 13:50:42
Job time: 190 sec